UNIVERSITY OF LONDON THESIS

Degree phd Year 2007 Name of Author TAMARA CHARLOTTE LORRAINE CAVANNA

Microarray Expression Analysis of Metastasising Sarcoma Cells Implies a Role for Protein 4.1B in Metastasis

Tamara Charlotte Lorraine Cavanna

A thesis submitted in partial fulfilment of the requirements of the degree of

Doctor of Philosophy

University of London

August 2006

ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

This thesis was completed under the supervision of Dr. Daniel Zicha.

Thesis Committee members were
Prof. Anne Ridley (second supervisor) and Dr. Michael Way.

The work described here was carried out at the Cancer Research UK London Research Institute, 44 Lincoln's Inn Fields, London WC2A 3PX.

I, Tamara Charlotte Lorraine Cavanna, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Abstract

The identification of the genes underlying the mechanisms of metastasis is of great interest, since metastatic disease is the major cause of death in cancer patients. To investigate the possible mechanisms of metastasis, I used an inbred rat sarcoma model and analysed four related cell populations with significant differences in their ability to shed metastases when subcutaneously injected into inbred rats ($P < 0.01$). I characterised the motility and morphology of the four cell populations and found that the metastatic cells had a stronger chemotactic response to PDGF/IGF, migrated faster, and had fewer stress fibres, than the non-metastastic cells.

Next I performed microarray analysis to investigate the gene expression differences underlying the progression to the metastatic phenotype. I examined gene expression in the cultured cells, and in the tumours that formed after subcutaneous injection of the cells into rats. In this way, I was able to identify genes whose expression was significantly changed with metastatic potential, both in cultured cells and in primary tumours. Twenty-three genes were differentially expressed more than 2.5-fold in metastatic cells ($P < 0.05$).

The gene encoding protein 4.1B was down-regulated in the metastatic cells. To investigate the possible function of 4.1B, I reduced its expression in non-metastatic cells by RNA interference (RNAi). Cells with reduced 4.1B expression displayed an altered F-actin morphology, with a loss of stress fibres compared to control cells. The stress fibre phenotype was rescued by transfection of an RNAi-resistant 4.1B cDNA, showing that the effect was specific. I also found that the 4.1B RNAi cells migrated at twice the speed of the wild-type cells.

I conclude that the loss of 4.1B in the metastatic cells causes a significant loss of actin stress fibres and increase in cell speed, and thus plays a role in progression to metastatic phenotype.

Acknowledgements

During my PhD I have been supported by a vast and diverse network of people, just a few of whom are mentioned here. First of all, I would like to thank my supervisor Daniel Zicha for giving me the opportunity to work in his laboratory. It has been a priviledge to work in the Light Microscopy Laboratory and to be exposed to many different techniques under the expert guidance of Daniel, and the members of his lab, Deborah Aubyn, Colin Gray, Yan Gu, Peter Jordan and Alastair Nicol. I was also lucky enough to have Anne Ridley and Michael Way as my thesis committee, and I thank them for their direction and understanding over the years.

Microscopy is lovely but sometimes needs to be diluted with unpalatable things such as cloning and immunoblotting. Thank you to the many friendly and clever people at the London Research Institute who have helped me make the most of those techniques. Your advice, always generously given, is greatly appreciated. I am particularly grateful to Susanne Adams, Fabrizia Cesca, Joanne Durgan, Anna Hebden and Tan Choon Ping. Thanks to Ian Morris for always making sure I had everything I needed, and to the Equipment Park for their friendly efficiency.

I also acknowledge the help of Peter Parker's lab, whose weekly meetings I attended. Thank you for not hurting my feelings regarding the appalling western blots I showed you over the last three years, and for all your helpful suggestions. I am also indebted to Angus Cameron, Fabrizia Cesca, Tan Choon Ping, Erik Sahai, Christine Schmidt and Daniel Worth for taking time out of their schedules to read and critically comment on my thesis.

When spending the majority of one's waking hours reflecting on how fast cells crawl, and why, one rather begins to lose one's mind. Thanks to all my wonderful friends for always being there and keeping me sane; too many to mention, but you know who you are and how grateful I am.

Last but by no means least, thanks to my parents for providing some pretty nice DNA to get started with, and to my whole family, just for being them.

For my parents

Table of Contents

List of Figures

List of Tables

List of Appendices

Abbreviations

A297 A297Nb cell

A311 A337/311RP cell

AcGFP Green Fluorescent Protein, a new variant derived from *Aequorea coerulescens*

ADP adenosine diphosphate

ATP adenosine triphosphae

Bk Brain and Kidney Protein

BS Bovine Serum

CASK Calcium/Calmodulin Serine Kinase

cDNA complementary DNA

CFP Cyan Fluorescent Protein

CMV Cytomegalovirus (promoter)

cRNA complementary RNA

DAL-1 downregulated in adenocarcinoma of the lung

DCIS Ductal carcinoma in situ

DEPC Diethylpyrocarbonate

DIP Database of Interacting Proteins

DMSO Dimethylsulphoxide

DNA Deoxyribonucleic acid

dNTP deoxynucleotide

DTT Dithiothreitol, threo-2,3-dihydroxy-1,4-dithiolbutane

ECM Extracellular Matrix

Epb4113 erythrocyte protein band 4.1-like 3

ER Expression ratio

EST Expressed Sequence Tag

FA Formaldehyde

F-actin Filamentous actin

FERM 4.1, Ezrin, Radixin, Moesin (domain)

FITC Fluorescein isothiocyanate

FLAP Fluorescence Localisation After Photobleaching

GFP Green Fluorescent Protein

GO Gene Ontology

IF Intermediate Filament

IGF Insulin-like growth factor

K2 LW13K2 cell

Kb kilobase

LA Luria's Agar

LB Luria's Broth

MEM Minimum Essential Medium

MLC Myosin Light Chain

MLK3 Mixed Lineage Kinase 3

mM millimolar

nm nanometre

nM nanomolar

NSCLC non-small cell lung cancer

PBS Phosphate-buffered saline

PCR Polymerase Chain Reaction

PDGF Platelet-derived growth factor

PDGFR PDGF receptor

Pmol picomol

PRMT3 protein N-methyltransferase 3

PVDF polyvinylidene difluoride

RCC Renal cell carcinoma

RCCC Renal clear cell carcinoma

ROCK, ROK Rho kinase

Rpm revolutions per minute

RNA ribonucleic acid

RNAi RNA interference

RT-PCR Reverse Transcription Polymerase Chain Reaction

SDS-PAGE SDS polyacrylamide gel electrophoresis

siRNA short interfering RNA

T15 RPSL4T15 cell

TBE Tris Borate EDTA buffer

TBST Tris-Buffered Saline/Tween-20

µg microgram

µl microlitre

µM micromolar

µM micrometre

UV ultraviolet

YFP Yellow Fluorescent Protein

Chapter One

Introduction

1.1 Metastatic disease

1.1.1 Cancer and metastasis

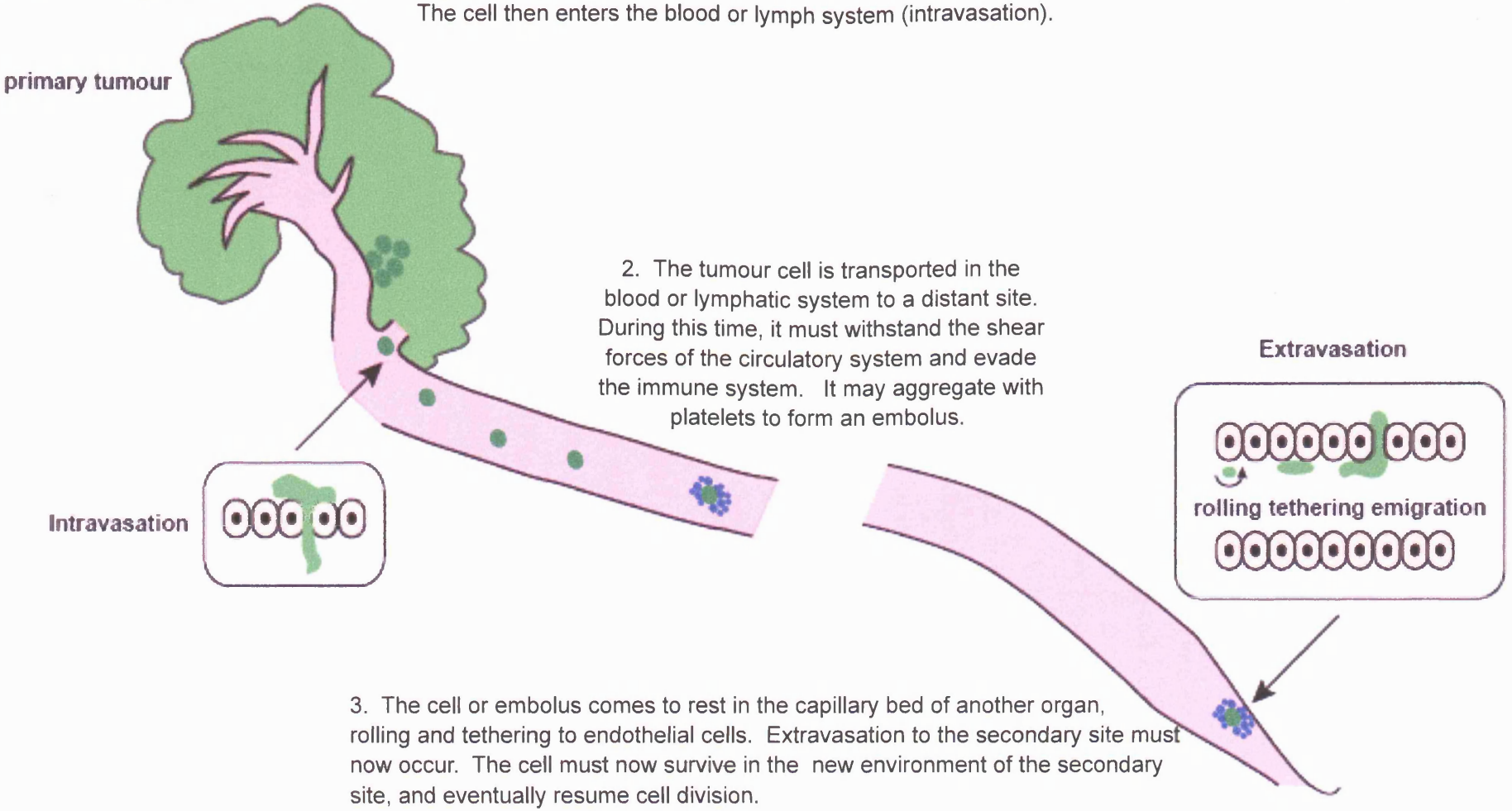
The uncontrolled division of cells leading to the formation of a tumour is known as cancer. Cancer is the result of the deregulation of a certain number of genes by a range of environmental and genetic factors (Hahn and Weinberg, 2002). The most life-threatening aspect of cancer is the ability of cancer cells to acquire the ability to migrate through the body to form one or more secondary tumours. This process is called metastasis, and it accounts for the majority of deaths in cancer patients (Ahmad and Hart, 1997). The development of metastases is insidious and often underway even before the primary tumour has been detected. Therefore it is important to understand the cell behaviour and associated gene expression of metastatic tumours, with a view to improving treatments and reducing the incidence of metastasis in cancer patients.

1.1.2 The metastatic cascade

To metastasise, the tumour cell needs to progress through a series of critical steps known as the metastatic cascade (Ahmad and Hart, 1997). The metastatic cascade, summarised in Figure 1.1, begins when an embolus of cells or a solitary cell detaches from the primary tumour. Detachment may occur in a few ways, such as the inadvertent dissemination of cells during surgical resection of the tumour, or a reduction in the strength of cell-cell adhesions in the tumour (Bogenrieder and Herlyn, 2003). The tumour cell then enters the blood or lymphatic system; this can occur via vessels vascularising the tumour itself. Alternatively, the tumour may invade the local tissue before entering the circulation. Upon intravasation (entry into the vessel) the tumour cell must withstand the shear forces and other mechanical stresses of the circulatory system, and continue to evade the immune system. Whilst in the blood stream, it may acquire chaperones, for example platelets, which aggregate with the tumour cell and provide protection in the circulation (Karpatkin and Pearlstein, 1981). The metastasising cell then arrests in the capillary bed of another organ, and begins extravasation. This may occur in a passive way, by the tumour cell becoming lodged in a capillary bed through which it passes by attrition of the vessel wall. Alternatively, intravasation can occur in a more active way, involving the tumour cell degrading the ECM, and crossing the wall of the vessel by regulating its shape and motility. Once at the secondary site, the cell needs to survive in the new environment, and eventually resume cell division to form the secondary tumour. The specific behaviours that enhance the metastatic potential of a tumour cell are described in more detail in the following section.

Figure 1.1: The metastatic cascade.

1. The metastatic cascade begins when an embolus of cells or a solitary cell detaches from the primary tumour. The cell degrades the extracellular matrix of the surrounding tissue, with matrix-degrading proteases, which may come from the tumour cell or from other cells in the tumour microenvironment. The cell then enters the blood or lymph system (intravasation).



1.1.3 The events and cell behaviours required for metastasis

In metastasis, a subpopulation of cells from the primary tumour is selected for its ability to negotiate the metastatic cascade (Fidler and Hart, 1982). Cells in the metastatic cascade must overcome a series of barriers. Accordingly, if the tumour cell lacks the characteristics necessary to progress through any part of the metastatic cascade, it cannot cause metastases. Therefore, each of the behaviours required for the metastatic cascade can be studied with a view to being made a target for anti-metastatic therapies. The events and behaviours that might be needed by the metastasising cell are considered here.

1.1.3.1 Tumour vascularisation

Tumours more than 2 mm in diameter must be vascularised in order to survive (Fidler and Ellis, 1994). In fact, many adults carry microscopic colonies of cancer cells (in situ tumours) but never develop disease because of the failure of these tumours to recruit their own new blood supply (Black and Welch, 1993). However, when tumours do become vascularised, this provides an opportunity for cells to be transported to the rest of the body via the circulation.

1.1.3.2 Degradation of extracellular matrix

Degradation of extracellular matrix (ECM) can be involved in the local invasion at the start of metastasis, and entry and exit of the circulatory system. It is achieved by matrix-degrading proteases such as matrix metalloproteases (MMPs) (Ahmad and Hart, 1997). However, the proteases are not always directly made by the tumour. The tumour can influence neighbouring stromal cells, such as mast cells, fibroblasts and macrophages, to produce proteolytic enzymes. For example, Sato et al. (2004) have shown that contact between cervical carcinoma cells and peripheral stromal fibroblasts *in vitro* causes an increase in MMP production and activation in both cell types, and a subsequent increase in invasiveness of the carcinoma cells. Mast cell degranulation may also play a role in ECM degradation; the accumulation and degranulation of mast cells has been seen at the zone of tumour invasion (Kankkunen et al., 1997). Cells can contribute to ECM degradation by mechanical, as well as biochemical means. Tumour-associated macrophages have also been seen to form podosomes (Evans et al., 2003) which are associated with remodelling of the ECM, and may promote cancer cell invasion by remodelling ECM both within the tumour and at the tumour-stroma interface. Finally, Lacal et al. (2005) found that the cytokine vascular endothelial

growth factor-A (VEGF-A), expressed by endothelial cells, can act on human melanoma cells and cause them to produce VEGF-A, which increases their ability to invade the ECM.

1.1.3.3 Regulation of adhesion

Changes in adhesion are thought to be necessary for tumour cell metastasis at various stages of the metastatic cascade (Bogenrieder and Herlyn, 2003). Initially, cell adhesion in the solid tumour mass should be weak enough that one or more cells may be freed. In the blood stream or lymph system, the aggregation of a tumour cell with platelets may help it to easily lodge in capillaries and facilitate the tethering and adhesion to endothelial cells in order to exit the vessel (Karpatkin and Pearlstein, 1981). Here, it is necessary to express molecules that can interact with molecules on the surface of the endothelial cells, for example CD44 (Marhaba and Zoller, 2004). The tumour cell may arrest on the vessel wall by attaching to an endothelial cell, or to basement membrane exposed by endothelial cell retraction. Metastatic cells can also increase their chances of achieving transendothelial migration by stimulating endothelial cell retraction. Hart et al. (2005) showed that the binding of prostate epithelial cells to endothelial cells can cause significant endothelial retraction. The basement membrane exposed upon cell retraction is generally a better adhesive substrate for the tumour cells than the endothelial cell surface and can allow the tumour cell to move more easily through the endothelial barrier.

1.1.3.4 Cell motility

It is widely accepted that cell motility is essential for metastasis. EMT (epithelial to mesenchymal transition) is a hallmark of many cancers and results in the acquisition of motility (Thiery, 2003). Cell motility is important since the metastasising cell needs to move out of the primary tumour, pass between cells in the local tumour environment, and eventually transmigrate through vessel walls for entry and exit of the blood or lymphatic systems (Condeelis et al., 2005). Heightened motility is a common feature of metastatic cells and has been described by several groups. In fact, a recent *in vivo* study has shown that metastatic tumour cells move 4.5 times as frequently as non-metastatic tumour cells in the same tumour microenvironment (Sahai, 2005).

Motility of tumour cells may be directional, in the sense that organs and endothelial cells might secrete factors that can promote migration of cancer cells into the organ/vessel itself. Hujanen and Terranova (1985), for example, have shown that

melanoma, sarcoma and breast cancer cells can chemotax towards extracts from the organs to which they preferentially metastasise. *In vivo*, tumour cells may migrate to the secondary site by travelling along the homing pathways used by other cells, for example by exploiting the chemokine-mediated homing of haematopoietic cells to specific sites in the body. Muller et al. (2001) report that the chemokines CXCR4 and CCR7 are associated with the metastasis of melanoma and breast cancer and propose that chemokines and their receptors promote metastasis of cancer cells.

1.2 Cell motility and the cytoskeleton

1.2.1 Cell motility and the cytoskeleton in health and disease

Cell motility underpins a number of physiological processes, as well as metastasis. The first cell movements occur during embryogenesis, for example, migration of the neural crest cells or the migration of growth cones during the establishment of the nervous system (Bray, 2001). Cell migration remains important throughout adult life, for example in the migration of fibroblasts during wound healing, and in the circulation of immune cells surveying for infection (Bray, 2001). In order to migrate, cells need to modulate their shape and their interactions with the substrate. Cells are able to do this by altering their cytoskeleton, which is composed of three types of protein filament; intermediate filaments, microtubules and actin.

1.2.2 Intermediate filaments and microtubules

Intermediate filaments (IF) provide the cell with mechanical strength and perform a variety of structural functions. They are comprised of a number of proteins, for example cytokeratins, vimentin, desmin and nuclear lamins (Alberts, 2002), each of which is specialised for a different function, for example, the lamins make a meshwork underlying the nuclear membrane, called the nuclear lamina, whilst other types of IF extend into the cytoplasm from the nuclear lamina or link cells at gap junctions. IF are predominantly found in cells that require a lot of mechanical strength, such as epithelial cells (Alberts, 2002).

Microtubules are long, thick polymers of tubulin, found in the cytoplasm. One end of the filament is usually attached to the centrosome, close to the nucleus, whilst the other end (the plus end) is capable of rapid growth, reaching the edge of the cell. Microtubules are highly dynamic structures which are also capable of rapid

depolymerisation and as such are said to possess dynamic instability (Waterman-Storer and Salmon, 1997). Microtubule-associated proteins (MAPs) can stabilise microtubules or modify their function, for example, motor proteins such as kinesin and dynein allow the movement of organelles along microtubules by binding to and trafficking organelles along microtubules (Alberts, 2002). Microtubules can play a role in cancer, and some anticancer drugs, for example Taxol, target microtubules (Jordan and Wilson, 2004).

1.2.3 Actin

Whilst intermediate filaments and microtubules have a role to play in cell motility and cell shape, actin is of particular importance in this thesis; hence it is described in more detail here. Actin is an abundant protein that can exist as a globular protein (G-actin) or as Filamentous actin (F-actin). Actin polymerisation can generate forces in the cell (Miyata and Hotani, 1992) which are used to create dynamic structures that co-ordinately result in cell motility.

1.2.3.1 Actin polymerisation

Actin polymerisation begins with the formation of a stable actin trimer. Two actin monomers bind relatively weakly to each other and therefore are rather unstable, but the addition of a third actin monomer makes a stable group. This is called nucleation, and it is the rate-limiting step in actin polymerisation (Bray, 2001). The monomers add to each other in the same orientation, so that filaments are composed of uniformly oriented molecules. The ends of the filament have different characteristics; a fast-growing plus (barbed) and a slow-growing minus (pointed) end. Each actin monomer is tightly associated with an ATP molecule which is hydrolysed after polymerisation occurs. The dissociation of the phosphate group reduces the affinity of the actin molecule for its neighbour and hence promotes depolymerisation, meaning that actin filaments are dynamic structures (Bray, 2001). Actin polymerisation is regulated by actin binding proteins, examples of which are cofilin, profilin and thymosin. Profilin binds actin monomers (Tilney, 1978) and enables their transfer to the plus end of filaments. Cofilin causes the disassembly of actin monomers from the pointed end of the actin filaments (Lappalainen and Drubin, 1997). Thymosin acts as a buffer of actin monomers by sequestering them and releasing them into the cytosol when the rapid formation of F-actin is required (Pollard et al., 2000).

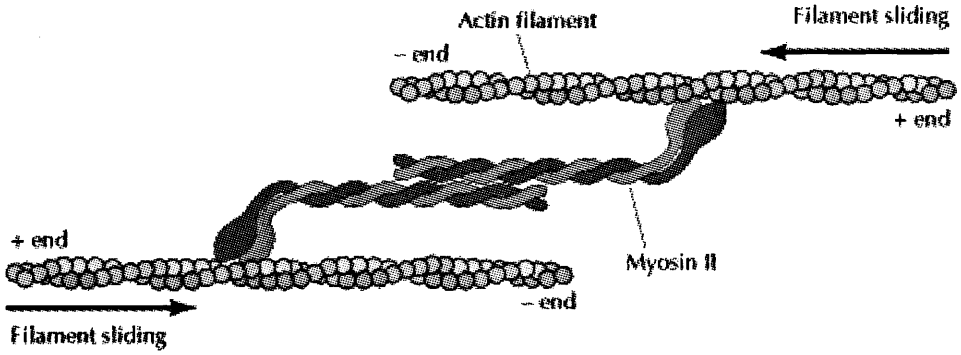
1.2.3.2 Actin treadmilling

Actin filaments are in a state where the continual addition of subunits at one end of the filament is matched by loss of subunits at the other end; this is known as actin treadmilling (Alberts, 2002). Treadmilling allows the filament to remain the same length as individual actin subunits move along it. Actin treadmilling is not equilibrium behaviour, and consequently depends on the energy released by the ATP hydrolysis which occurs upon actin polymerisation (Alberts, 2002). The advantage of this dynamic behaviour is that the cell can alter its shape rapidly in response to stimuli.

1.2.3.3 Actomyosin contraction

Actomyosin is a complex of actin and myosin. Actomyosin contraction is responsible for the movements in muscle fibres as well in non-muscle cells. Myosin filaments can interact with actin filaments and generate movement by the hydrolysis of ATP, where two actin filaments are pulled past each other in opposite directions, causing contraction (Lodish, 1999). This is illustrated in Figure 1.2.

Figure 1.2: Actomyosin contraction in non-muscle cells.



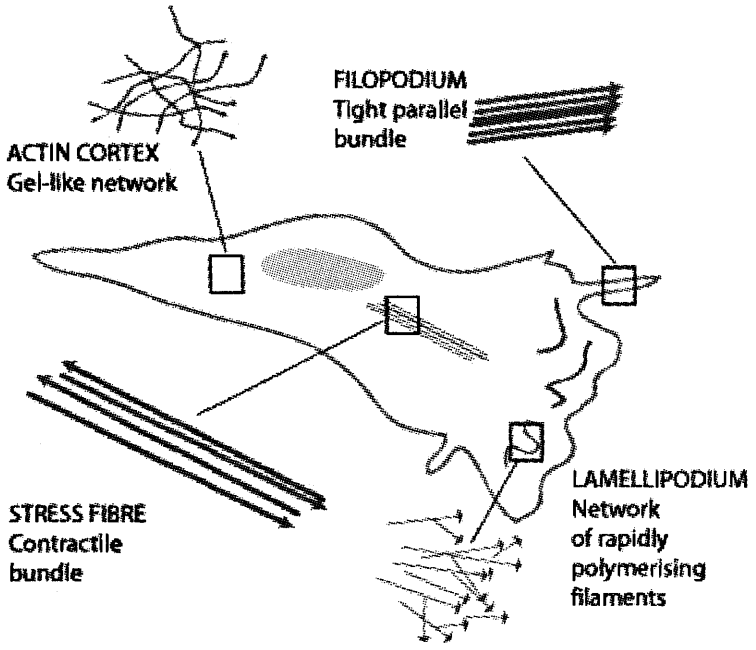
In actomyosin contraction, tilting of the head domain of myosin causes contraction as the two actin filaments slide past each other. This is taken from a diagram in Molecular Cell Biology (Lodish, 1999).

Actomyosin contractions are involved in cytokinesis, where the cytoplasm of a cell divides following nuclear division, and in migration. Actomyosin is the basis of stress fibres, which will be introduced shortly. Actomyosin contraction is the result of a signalling cascade mediated by the small GTPase Rho. Briefly, Rho activates Rho kinases (ROCKs) which in turn phosphorylate the myosin light chain, either directly or indirectly, causing contraction (Katoh et al., 2001). Myosin light chain phosphorylation can also be brought about by the activity of other proteins for example ZIP-like kinase/hZIPK (Endo et al., 2004; MacDonald et al., 2001).

1.2.4 The actin cytoskeleton

Actin can form a variety of specialised cellular structures, with the help of a wide range of actin binding proteins. Actin binding proteins can produce actin networks with distinct qualities, for example, α-actinin cross-links actin filaments into bundles, and spectrin forms gel-like networks (Bray, 2001). Examples of actin-based structures, stress fibres, lamellipodia, filopodia and the actin cortex, are displayed in Figure 1.3.

Figure 1.3: Assemblies of actin filaments in a fibroblast.

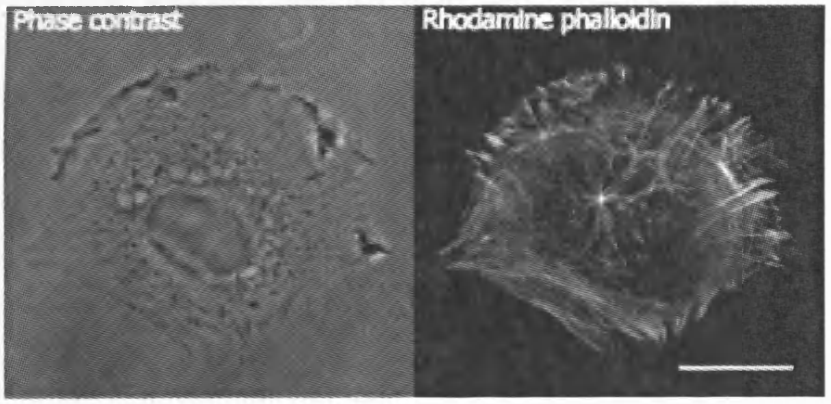


Actin can be arranged into different structures specialised for the various requirements of the cell. In fibroblasts, actin is arranged into a gel-like network called the actin cortex, which supports the cell membrane. It can also occur in tight parallel bundles in plasma membrane extensions known as filopodia. At the leading edge of a migrating cell, branched actin filaments in the lamellipodium polymerise to push the cell forward. Actin can also be incorporated into contractile bundles called actin stress fibres, to give the cell contractility and structural integrity. This schematic is adapted from a figure in Molecular Biology of the Cell (Alberts, 2002).

1.2.4.1 Stress fibres

Stress fibres, shown in the phalloidin-stained fibroblast in Figure 1.4, are bundles of actin, myosin and other accessory proteins. Stress fibres are long, plasma membrane-bound cables which can stretch from one side of the cell to the other, and exert tension by contracting. One of the actin binding proteins involved in stress fibres, α-actinin, becomes relevant later in this thesis and so will be briefly described here. α -Actinin localises along stress fibres and is thought to play a role in stress fibre bundling as well as linking stress fibres with the ECM. It exists as a dimer, and contains two actin binding domains which cross-link actin filaments, holding them approximately 40 nm apart (Alberts et al., 2002). The reason that the fibres are held apart rather than closely apposed is that myosin II, the protein responsible for stress fibre contractility, needs to access the bundle. α-Actinin may also be involved in responses to IGF in some cells (Guvakova et al., 2002), and it has also been observed in invadopodia (Chen, 1989), structures exhibited by cells migrating in a three dimensional matrix, and may therefore have an important role in cell motility.

Figure 1.4: Actin stress fibres in a fibroblast.

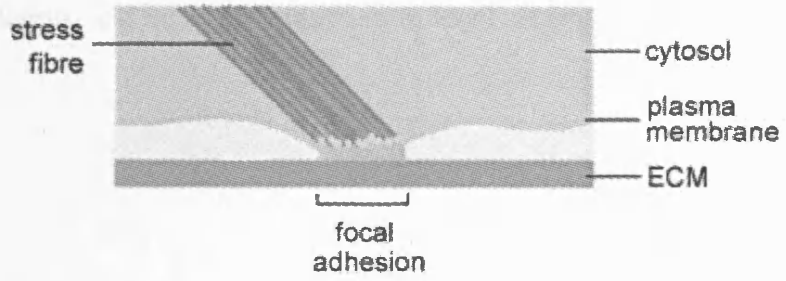


This fibroblast was fixed and stained with a fluorescent conjugate of the actin binding protein phalloidin. This confocal image shows actin stress fibres spanning the rear of the cell. Scale bar = 20 μm.

1.2.4.2 Focal adhesions

Stress fibres are embedded into structures known as focal adhesions, which are clusters of proteins forming an attachment between the cell and the substrate. A schematic of a stress fibre meeting a focal adhesion is shown in Figure 1.5.

Figure 1.5: Schematic cross-section of a stress fibre meeting a focal adhesion.



This diagram illustrates a stress fibre attaching to the cell membrane at a focal adhesion. The focal adhesion is formed by the clustering of integral membrane proteins which transduce signals from the ECM to the cell. The stress fibre indirectly binds the ECM at the focal adhesion via transmembrane proteins such as integrins. This diagram is modified from one presented in Cell Movements (Bray, 2001).

Focal adhesions are composed of many different proteins. Integrins within the focal adhesions bind the ECM, whilst the cytoplasmic domain of the integrin engages with other focal adhesion proteins, which include α-actinin, vinculin, paxillin and src. As actin filaments can associate with the proteins bound to the integrin's intracellular domain, the integrin enables actin filaments to be linked to the ECM (Bray, 2001). In this way, focal adhesions link the cytoplasm with the substrate. As well as mechanically attaching cells to the substrate, focal adhesions also participate in intracellular communication, transmitting signals for various signalling pathways and acting as signalling complexes for the cell. Focal adhesions are required for cell motility in many cell types including fibroblasts (see Figure 1.9) which need focal adhesions to be formed at the leading edge of the migrating cell. These adhesions then remain fixed (relative to the substrate) as the cell moves forward, and are disassembled at the rear of the cell.

How do focal adhesions form? Cells grown on an untreated substrate in the absence of serum do not naturally form focal adhesions. Focal adhesions form during spreading or migration on a tissue culture surface providing the substrate contains the correct ECM

components (Sastry and Burridge, 2000). Inititally, focal complexes form as the cell contacts the substrate and these complexes then mature into focal adhesions which are more stable attachment sites which also anchor actin filaments. Another requirement for focal adhesion formation is cell contractility or tension, and this is produced by myosin activity in the stress fibres. The small GTPase Rho, through its downstream effector Rho kinase, stimulates myosin II chain phosphorylation and hence increases actomyosin contractility. This results in the formation of stress fibres and the clustering of integrins and other proteins to form focal adhesions (Sastry and Burridge, 2000).

1.2.4.3 The actin cortex

The actin cortex is a dense network of actin filaments tightly associated with the cell membrane, as illustrated in Figure 1.6. Actin filaments are attached to the cell membrane via adapter proteins (such as α-Actinin) which bind both actin and integral membrane proteins (Bray, 2001).

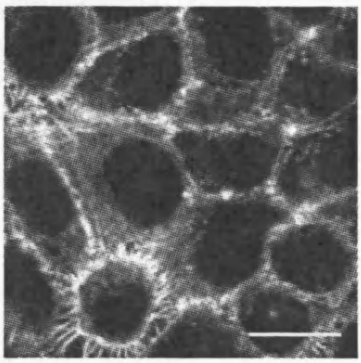


Figure 1.6: Cortical actin visualised by staining with rhodamine phalloidin.

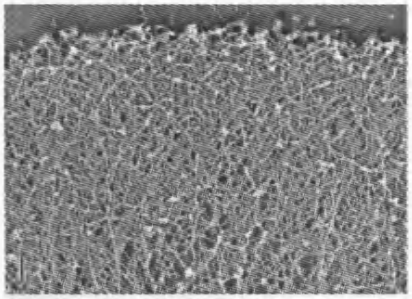
Hela cells were fixed and stained with rhodamine phalloidin, and imaged on a confocal microscope. This image, of a section of the cell near the apical surface, shows the cortical actin underlying the cell membrane. Scale bar = 20 μm.

The nucleation and polymerisation of cortical actin produces protrusions on the cell surface, such as invadopodia, filopodia and lamellipodia.

1.2.4.4 Lamellipodia

Lamellipodia are broad, flat protrusions of the cell periphery lying parallel to the substrate. They typically represent the leading edge of the cell. When lamellipodia sweep back over the cell as it migrates forward, they appear as ruffles, examples of which can be seen on the migrating cell in Figure 1.8. They are composed of a meshwork of branched actin filaments, with the fast-growing ends facing outwards (Svitkina and Borisy, 1999). The cell rapidly polymerises actin at its leading edge causing a lamellipodium to extend across the substrate in the direction of migration. The branching of actin at the lamellipodium generates more force than the simple polymerisation of one, unbranched filament, and allows the membrane to be pushed forward rapidly. Figure 1.7 shows the branched actin filaments in a fibroblast lamellipodium. The branching is mediated by the Arp2/3 complex binding the F-actin filament and nucleating a new filament (Svitkina and Borisy, 1999). Actin polymerisation at the leading edge and indeed in other actin-based protrusions relies upon the continual delivery of actin monomers to the end of the growing filament. The speed of actin delivery to the leading edge of the cell was recently calculated after FLAP analysis of sarcoma cells (Zicha et al., 2003) and found to be greater than that expected due to diffusion alone, meaning that other factors are likely involved in the delivery of actin monomers to the growing ends of filaments.

Figure 1.7: Branched actin networks in a lamellipodium.



This scanning electron micrograph shows the actin filaments in the cytoskeleton of a cultured fibroblast (REF-52). Binding of the Arp2/3 complex onto actin filaments allows nucleation of actin monomers leading to the polymerisation of a branch. Modified from a micrograph by T M Svitkina (Svitkina et al., 1997).

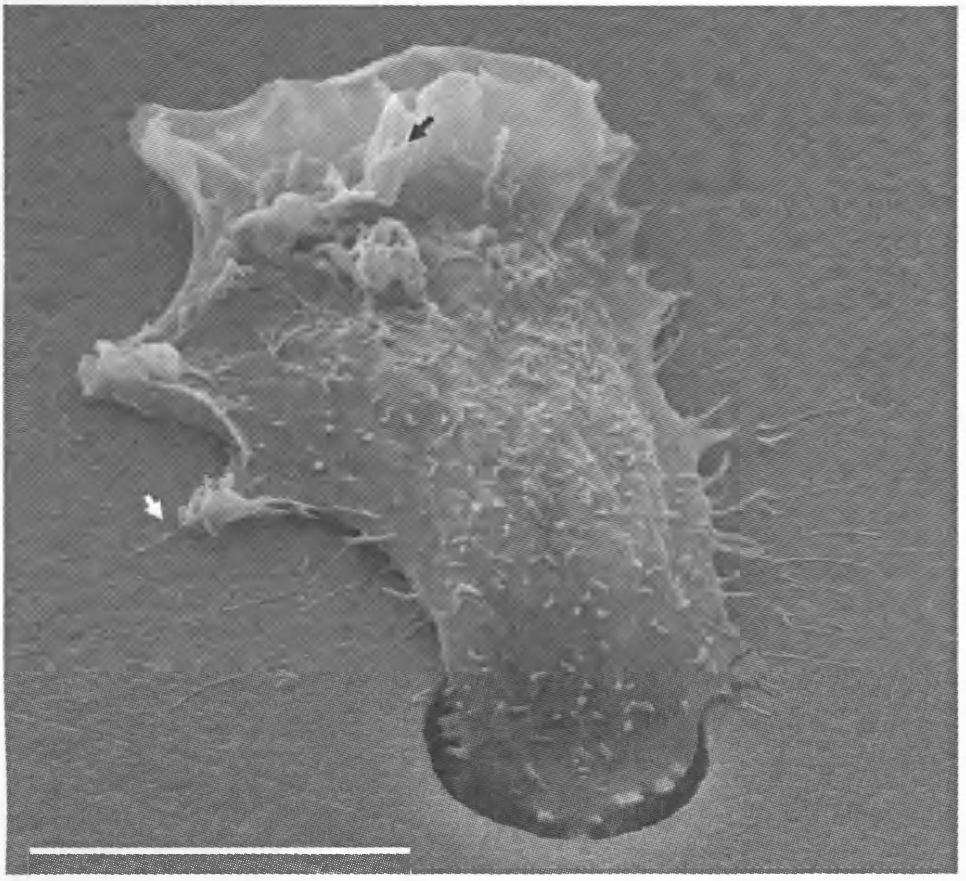
1.2.4.5 Filopodia

Filopodia are active membrane protrusions, approximately 0.15 μm in diameter, containing unipolar actin filaments (Small et al., 2002). They are found in a variety of cell types, including fibroblasts and growth cones, and allow the cell to explore its environment. The extensions at the sides of the cell in Figure 1.8 may be filopodia being used to sense the environment as it migrates.

1.2.4.6 Podosomes and invadopodia

These are actin-based structures which have only recently started to receive widespread attention. Podosomes appear by electron microscopy as small (1-2 μm diameter, 200-400 nm deep) plasma membrane extensions, whereas invadopodia are much larger (8 μm wide, 2 μm deep) (Buccione et al., 2004). Podosomes are actin-rich extensions of the cell membrane which may function as dynamic adhesion structures for cell migration. They are distinct from focal adhesions as they contain proteins required for actin polymerisation, for example WASP (Calle et al., 2004). Invadopodia are actin-rich and contain other proteins such as Vinculin and α-Actinin (Chen, 1989). They can also degrade the ECM by secreting matrix-degrading enzymes, and they are seen in invasive cancer cells cultured on a physiological substrate and may help clear the path for a cell as it migrates.

Figure 1.8: Scanning electron micrograph of a migrating fibroblast.



The cell in this scanning electron micrograph is migrating through the 8 µm diameter pore of a polyethylene terephthalate membrane. This image illustrates the extensive and complex structures arising from the cell membrane during migration. The lamellipodium has extended and formed attachments with the substrate, allowing the cell to pull itself forward (indicated by white arrow). As the cell migrates, the large lamellipodium is swept backwards over the cell, forming ruffles (indicated by black arrow). The thin filaments visible on both sides of the cell are filopodia or retraction fibres. Dorsal structures are also clearly present. These membrane structures are all due solely to the remodelling of the actin cortex. This micrograph was taken by Steve Gschmeissener. Scale bar = 10 µm.

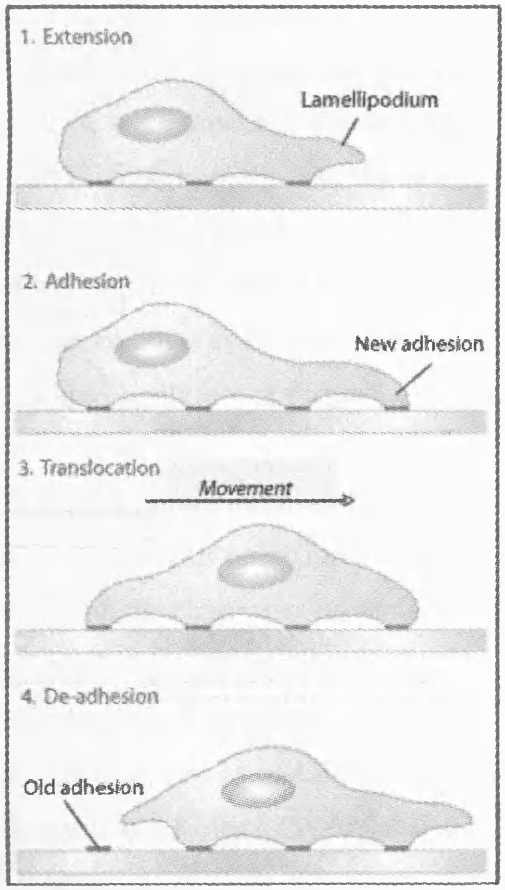
1.2.5 Fibroblast migration

Cell migration involves the coordinated remodelling of the actin cytoskeleton into the actin-based structures described in Section 1.2.4. I will introduce fibroblast migration since the model used in this study is fibroblast-like, although it should be noted that mechanisms of migration vary greatly according to cell type.

1.2.5.1 The multi-step model of fibroblast migration

Fibroblast migration relies upon a well defined repertoire of movements; extension of a lamellipodium, adhesion of the extended membrane to the substrate, contractile forces to shift the body of the cell forward in the direction of the lamellipodium, and detachment of the adhesions at the rear of the cell (Abercrombie and Heaysman, 1966). This is summarised in Figure 1.9.

Figure 1.9: Repertoire of movements needed for fibroblast migration.



The polarised cell extends the membrane at the front of the cell to make a lamellipodium (1), which is pushed forward over the surface by the rapid extension of actin filaments. New adhesions form between the lamellipodium and the substrate (2). The cell body shifts in the direction of the new membrane extension (3) by exerting contracting forces to move the cytoplasm forwards. The cell also needs to retract its tail and move it forward in the direction of migration (4). This requires the disassembly of the focal adhesions at the rear of the cell to allow the membrane to move away from the substrate. Note that these movements do not necessarily occur in the order depicted in the diagram. After Molecular Cell Biology (Lodish, 1999).

Having extended a lamellipodium, the cell now forms focal contacts with the substrate. Focal contacts are built around integrins, which are transmembrane proteins. The extracellular domain of the integrin binds to extracellular molecules such as fibronectin, and the cytoplasmic domain binds to proteins localised to the focal adhesion, for example α-Actinin or Vinculin (Bray, 2001). Figure 1.5 shows a schematic section through a focal adhesion, where a stress fibre contacts the substrate indirectly via integral membrane proteins. Contractile forces, generated through stress fibres, move the cell body forward to follow its lamellipodium. As essential as the formation of focal adhesions is that the cell releases the focal adhesions at the rear of the cell as it moves forward. This occurs partly through the exertion of tension on the rear focal adhesions, via actomyosin contraction (Katoh et al., 2001).

1.2.5.2 Cell migration in three dimensions

The study of cell migration has predominantly been carried out in two-dimensional (2D) models, where cells crawl on a topographically simple substrate such as plastic or glass. *In vivo*, cells are exposed to more challenging environments, where they need to migrate through rather dense and complicated tissues. Niggemann et al. (2004) investigated how the multi-step, two-dimensional model of cell migration (described in Section 1.2.5) compares to the motility of tumour cells in a three dimensional collagen matrix. They found that the four steps can be observed, although they often merge with each other, and do not always occur in the same sequence. Also, the first extension may not be carried out only by a lamellipodium but by a number of different pseudopodia, for example, a mixture of filopodia, invadopodia and podosomes.

Though the basic mechanisms of cell motility may be the same in two and three dimensions, cells likely have to operate slightly different modes of motility in order to deal with the challenges presented by the two environments. Indeed, Sahai and Marshall (2002) suggest that tumour cells can switch between an elongated and amoeboid mode of motility. This and other studies have suggested that two dimensional and three dimensional analyses reveal different aspects of cell motility (Martins and Kolega, 2006; Sahai and Marshall, 2002).

1.3 Protein 4.1B

1.3.1 FERM domain proteins

Whilst a large number of proteins are involved in the formation of stress fibres and in general cytoskeletal regulation, the FERM (4.1, Ezrin, Radixin, Moesin) domain-containing proteins will be discussed here since one member of this protein family is relevant to this thesis. The FERM domain has been shown to be involved in the linkage of the cytoskeleton to the ECM (Chishti et al., 1998). The best characterised FERM domain proteins are Ezrin, Radixin and Moesin. Ezrin was identified as an organiser of the cytoskeleton and membrane (Gould et al., 1989), radixin as a barbed end actin capping protein at the plasma membrane (Tsukita et al., 1989) and moesin as a cytoskeleton and plasma membrane linker (Lankes and Furthmayr, 1991). They are required for the induction of stress fibres and lamellipodia by the GTPases Rho and Rac, respectively (1997). Other FERM domain-containing proteins include NF2 and FAK, both of which are believed to play a role in linking actin to the ECM (McClatchey and Giovannini, 2005). Less well characterised are the 4.1 superfamily proteins. The first member of the family to be identified was 4.1R, as a major component of erythrocyte membranes (Conboy et al., 1986). Since then, the related proteins 4.1N, 4.1G and 4.1B have been found (Parra et al., 2000; Parra et al., 1998; Walensky et al., 1999).

1.3.2 DAL-1 and protein 4.1B

4.1B was simultaneously identified in a differential display screen of human primary non-small cell lung cancer as the gene DAL-1 which was ***D***ownregulated in ***A***denocarcinoma of the ***L***ung (Tran et al., 1999). DAL-1 was later found to be a functionally-complete fragment of the full length protein 4.1B (Gutmann et al., 2000; Gutmann et al., 2001). DAL-1 is functionally comparable to 4.1B in that it contains all the amino acids necessary for the tumour suppressor effect of 4.1B. 4.1B was identified in mice by Parra et al. and is so-called because of its membership of the 4.1 protein family and its high expression level in the brain (Parra et al., 2000). However it is now known to be expressed in a number of other tissues besides brain. In this thesis, the name 4.1B will be used since it relates the protein to the 4.1 family. In the description of studies where the functionally comparable fragment DAL-1 was used, I will refer to the protein as DAL-1/4.1B. Protein 4.1B is encoded by the gene *epb41l3*, which has

been found to encode a number of tissue-specific protein isoforms (Gascard et al., 2004; McCarty et al., 2005) via alternative splicing.

1.3.2.1 4.1B function

Like other members of the 4.1 family, 4.1B is thought to be involved in tethering the actin cytoskeleton to membrane glycoproteins. It is reported to be localised to the plasma membrane (Gutmann et al., 2000; Parra et al., 2000). It has been shown to have tumour suppressor effects by three different groups. Tran et al. (1999) introduced DAL-1/4.1B into non-expressing cell lines, and observed a significant reduction in cell growth. Charboneau et al. (2002) also found a growth suppressing effect in MCF7 breast carcinomas transiently transfected with DAL-1/4.1B, and Gutmann et al. (2000) found that it was lost in 60% of meningioma samples. However, they found no loss of 4.1B in sporadic schwannomas (out of 20 samples examined) which indicates that 4.1B may act in a tissue-specific manner.

In addition to the association of 4.1B loss with brain, breast and lung tumours, its tumour suppressor activity is supported by the fact that the chromosome region on which *epb41l3* is located, 18p11.3, is commonly deleted in tumours (Kittiniyom et al., 2001; Tran et al., 1998). 4.1B may also be involved in cellular attachment, since the expression of DAL-1/4.1B in MCF7 increases their adherence to substrates (Charboneau et al., 2002) and the expression of 4.1B/DAL-1 in schwannoma cells impaired cell spreading and migration in the Boyden chamber assay (Gutmann et al., 2001).

1.3.2.2 4.1B structure and its interactors

The structure of 4.1B (Parra et al., 2000) is outlined in Figure 1.10.

Figure 1.10: The domain organisation of 4.1B.

4.1B has an N-terminal FERM domain, which appears to be a membrane binding module common to a number of different proteins. In the cytoplasm, the FERM domain is in an inactive, closed conformation (Bretscher et al., 2002) and, upon activation (in the case of ezrin, by threonine/tyrosine phosphorylation) assumes an open conformation and moves to the membrane. The FERM domain of 4.1B is 74% homologous to the prototypical 4.1 family member, 4.1R. The FERM domain of 4.1R has been found to interact with various proteins including calmodulin and the integral membrane proteins glycophorin C and CD44; 4.1B is also likely to interact with these.

4.1B has been shown experimentally to interact with βII-spectrin (Gutmann et al., 2001) and the cytoplasmic tail of β8 integrin in primary astrocytes (McCarty et al., 2005). It was also found to interact with 14-3-3 (Yu et al., 2002) though a later study (Robb et al., 2004) showed that disruption of this interaction had no effect on the growth-suppressing effect of 4.1B, indicating that either the 4.1B/14-3-3 interaction does not occur *in vivo,* or that other proteins are capable of transducing the growth-suppressing signals.

There is a spectrin/actin binding domain (SAB), which, in the prototypical family member 4.1R, cross-links spectrin/actin dimers with glycophorin. Despite the sequence homology between the SAB of 4.1R and the putative SAB of 4.1B, the latter has been reported to interact very weakly with actin (Gimm et al., 2002) or in fact, not at all (Gutmann et al., 2001). The C-terminal domain (CTD) is homologous to the CTD of 4.1R which has putative binding sites for NuMA (Nuclear Mitotic Apparatus protein) and the glycoprotein FKBP13. There are three unique domains, U1, U2 and U3; U1, with unknown function, is upstream of the FERM domain, U2 is a spacer between the FERM and SAB domain and the function, if any, of U3 is currently unclear.

1.4 Gene expression profiling of metastasis

Just as the interplay of tumour suppressor and tumour promoter genes causes a cell to become cancerous, a particular combination of genes is required for the cell to progress through each step of the metastatic cascade. The development of microarray technology has enabled the investigation of the gene expression in tumours, which allows us to identify genes that influence metastasis (Rew, 2001). This has important clinical implications; microarray analysis of metastasis has the potential to improve diagnostic and prognostic tools, tailor adjuvant therapy to each patient, and identify molecular targets for drug development.

1.4.1.1 The need for gene expression analysis in cancer diagnosis

An important aspect of cancer patient care is the accurate assessment of the stage and likely outcome of the disease. A wide range of methods are currently used for this. These include the detection of specific molecules (e.g. Prostate Specific Antigen or HER2 receptor), lymph node status, histological grade of the tumour MRI imaging, CT scanning, and the detection of metastases in the circulation. Whilst these indicators are helpful in predicting prognosis, and designing treatment strategies, they are susceptible to failure (Dalton and Friend, 2006).

The inability to reliably predict cancer outcomes means that clinicians must err on the side of caution when planning treatment strategies for the patient. The result of this is that a large proportion of patients are significantly overtreated; patients typically receive adjuvant therapy (chemotherapy, radiotherapy or hormone therapy to kill cancer cells which may have begun to spread) after surgical resection of the tumour, although, not all patients will develop metastases anyway. Adjuvant therapies are exhausting and detrimental for the patient, and costly. Clearly it would be preferable to reliably predict which patients are at high risk of relapse and treat them accordingly, withholding unnecessary treatment from those whose risk of relapse is extremely low, and, in cases where more than one treatment option is available, selecting the treatment to which the cancer is most likely to respond.

The ideal way to make treatment decisions would be to characterise tumours on the basis of their molecular characteristics. For this reason it is hoped that gene expression analysis of tumours will reveal gene expression patterns- "predictive gene lists" or

"gene expression signatures"- that are associated with metastasis (van de Vijver et al., 2002).

1.4.2 Gene expression microarrays

A gene expression microarray consists of thousands of nucleic acid probes* bound in a specific pattern to a silicon or glass slide. Each probe is complementary to a transcribed DNA sequence from the organism of interest. Targets (nucleic acids from the sample) are produced by extracting RNA from a sample and converting it into cDNA. The cDNA target is then labelled and hybridised to the microarray, to which the probes are bound. Hybridisation between the sample and a probe indicates that a transcript complementary to the probe is being expressed by the cell. In this way, microarrays can measure the transcriptional activity in a sample. Gene expression microarrays are available as cDNA microarrays and oligonucleotide microarrays; I will focus on the latter as they were used in this study.

1.4.2.1 cDNA microarrays

cDNA microarrays are made by generating cDNA oligonucleotides by PCR, and printing them onto a solid substrate. cDNA microarray analysis involves the preparation of two samples to be hybridised to the same array (reviewed in Duggan et al., 1999). The targets are prepared by extracting RNA from the sample of interest, reverse transcribing it into cDNA, and labelling it with the fluorescent conjugate Cy3 or Cy5. A Cy3-labelled sample and a Cy5-labelled sample are competitively hybridised onto the same microarray, and unhybridised cDNA removed by gentle washing. Confocal laser scanning is used to detect the Cy3 and Cy5 signals on the microarray. The signal intensity of the Cy3 and Cy5 indicates the amount of hybridisation and thus the relative gene expression in the samples. Since two samples need to be hybridised together, selection of appropriate pairs of samples essential. For example, a tumour sample would need to be matched carefully to normal tissue or to a pooled tumour sample, in order to highlight genes which are up- or down-regulated in the tumour.

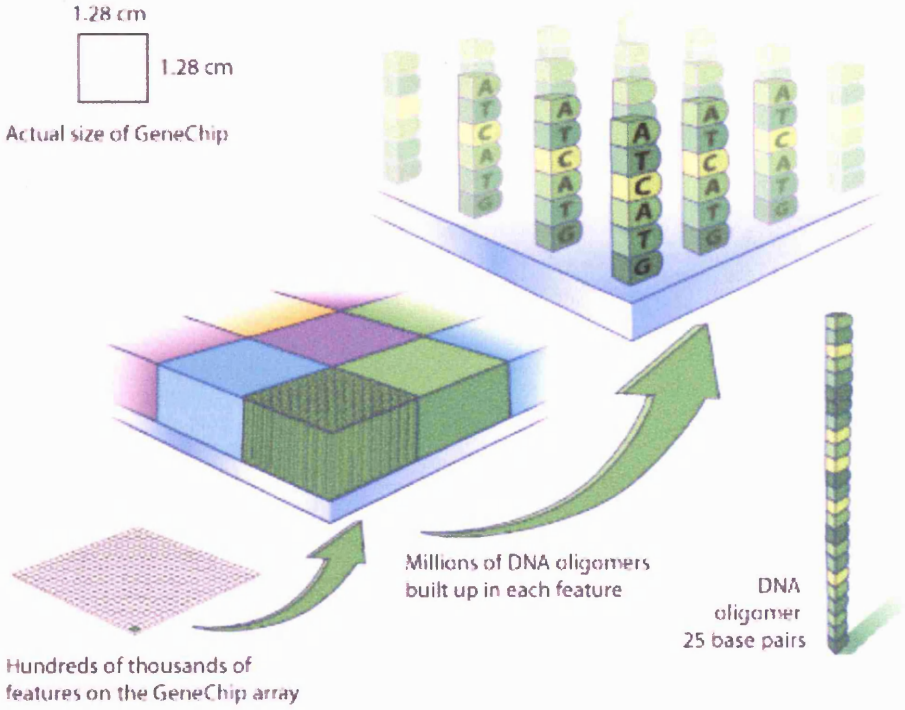
1.4.2.2 Oligonucleotide microarrays

Oligonucleotide microarrays are made by synthesising oligonucleotides directly onto the substrate by photolithography (light-directed synthesis), which enables accurate

* A note on nomenclature: in microarrays, the "probes" are the nucleic acid bound to the substrate, and the "targets" are the nucleic acids from the sample. This terminology may seem illogical given that in other hybridisation techniques, the probe is unbound. However it is recommended (Phimister, 1999) to define the probes and target not by whether or not they are immobilised but by the fact that the probe is a known entity that interrogates, or probes, the target (the unknown entity).

positioning of groups of oligonucleotides on the array (Lockhart et al., 1996). Currently, the most reliable and frequently used oligonucleotide microarrays are GeneChip® arrays, manufactured by Affymetrix, Inc (de Reynies et al., 2006). Each GeneChip array contains hundreds of thousands of different oligonucleotides which are present in millions of copies (Barone et al., 2001). The arrangement of features on an Affymetrix GeneChip® array is shown in Figure 1.11.

Figure 1.11: Schematic of features on an Affymetrix GeneChip® array.



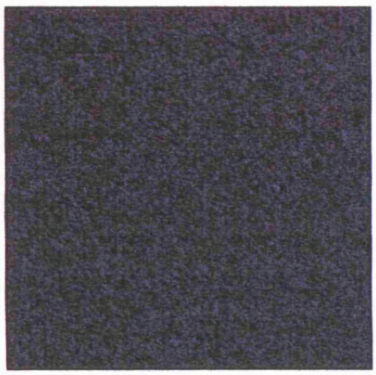
Each Affymetrix GeneChip® is 1.28 cm × 1.28 cm and contains hundreds of thousands of features. Each feature contains millions of 25 base-pair DNA oligomers probing for a particular sequence. Image courtesy of Affymetrix, Inc.

The Affymetrix system uses cRNA, rather than cDNA, targets to hybridsise to the probes. Targets are produced by extracting RNA from the sample, converting it to cDNA, and then performing *in vitro* transcription in the presence of biotinylated nucleotides to produce biotin-labelled cRNA. The cRNA is then fragmented and hybridised to the array, and unhybridised cRNA removed by gentle washing. The array is then incubated with the fluorescent molecule streptavidin-phycoerythrin, which binds to the biotin on the cRNA. The signal is amplified by an anti-streptavidin antibody followed by a biotin-labelled secondary antibody. The array is then scanned by a confocal laser, which excites the fluorescent molecules, causing them to emit light. Figure 1.12A shows an example image of a scanned GeneChip® array. The amount of fluorescence at each spot indicates the amount of hybridisation and thus the relative abundance of the transcript.

Figure 1.12 Example image of an Affymetrix GeneChip array and schematic explaining probe set design.

Images courtesy of Affymetrix, Inc.

(A) Scan image of full chip



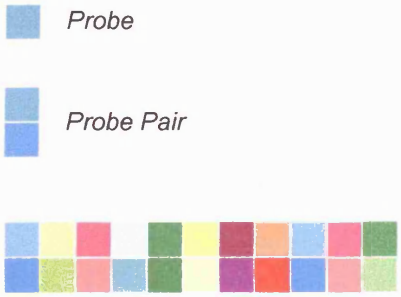
The GeneChip array contains

~ 330,000 Probes

~15,000 Probe Sets

~ 16,500 Probe Pairs

(B) Probes, Probe Pairs and Probe Sets

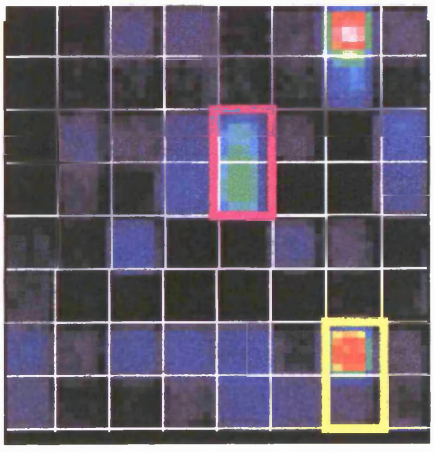


Probe

Probe Pair

Probe Set

(C) Close-up of scan image showing the PM/MM system



This close-up of a scan of a full chip shows 32 probe pairs, where the PM is the upper feature and the MM the lower feature of each pair.

The probe pair bounded by the pink box has a low PM/MM signal ratio, indicating non-specific hybridisation.

In contrast, the probe pair in the yellow box has a very high PM/MM ratio, indicating very little non-specific hybridisation.

1.4.2.3 Affymetrix probe design

Affymetrix GeneChip® arrays enable reliable measurement of transcript abundance because each probe on the array is actually paired in a probe set, with another probe that differs by one base. Each probe set (illustrated in Figure 1.12B) contains a Perfect Match (PM) probe, which has complete base pair complementarity to the transcribed sequence, and a Mismatch (MM) probe, which is the same as the PM probe with the exception of one mismatched base in the middle of the oligonucleotide. The MM probe is an indicator of non-specific hybridisation. After scanning, the fluorescence intensity from the MM probe is subtracted from the intensity of the PM probe in order to determine the intensity value for the particular probe pair. See Figure 1.12C for an example of the fluorescence intensities in the MM and PM of a probe pair. This probe design contributes to the accuracy of the gene expression analysis by reducing the likelihood of mistaking cross-hybridisation for perfect hybridisaton. A further advantage of the Affymetrix probe design is that each transcript is probed by 11 different probe pairs. Figure 1.12B illustrates the hierarchy of probe features on an Affymetrix GeneChip® array.

1.4.2.4 Overview of the two gene expression microarray platforms

The main differences between cDNA and oligonucleotide microarrays are summarised in Table 1.

Table 1: Differences between cDNA and oligonucleotide microarrays.

	cDNA microarray	Oligonucleotide microarray
Probe	PCR products printed onto the substrate	Oligonucleotides synthesised directly onto the substrate
Target	RNA extracted and converted into cDNA	RNA extracted, coverted into cDNA and then converted to cRNA
Labelling	Typically Cy3 or Cy5	Biotin
Number of samples hybridised	2	1
Hybridisation specificity	Moderate	High, owing to PM/MM probe design

Affymetrix GeneChip® arrays are a more accurate way of measuring gene expression, although cDNA arrays are certainly far more economical. The nomenclature commonly used for describing microarray experiments are recapitulated in Table 2.

Table 2: Summary of microarray nomenclature.

Microarray	A glass, filter, or silicon solid substrate, upon which oligonucleotides are deposited or synthesized in a defined spatial pattern.
cDNA microarray	A microarray produced by printing cDNA probes onto the substrate. Target cDNAs from two different samples are labelled with Cy3 or Cy5 and competitively hybridised to the microarray.
Oligonucleotide microarray	A microarray produced by synthesising the probes directly onto the substrate by photolithography.
Affymetrix GeneChip	A type of oligonucleotide microarray system in which multiple probes, including probes to measure non-specific hybridisation, measure the abundance of a transcript.
Target	An unknown DNA or RNA sequence in solution, labelled with a fluorescent conjugate such as Cy3, Cy5 or biotin.
Probe	A known, unlabelled DNA sequence tethered to the substrate.
Perfect Match Probe (PM)	A 25 base oligonucleotide with perfect sequence complementarity to the transcribed sequence.
Mismatch Probe (MM)	A 25 base oligonucleotide whose sequence is largely complementary to the transcribed sequence, apart from one mismatched base in the middle of the oligonucleotide. Hybridisation to the MM is used as a reference signal for non-specific hybridisation; a high PM/MM signal ratio indicates low levels of non-specific hybridisation.
Probe Pair	Comprised of a PM and MM probe.
Probe set	A series of probe pairs probing for the same transcript.
Signal	The fluorescence intensity at a location on the array, indicating the abundance of a transcript.
Expression Ratio (ER) or Fold Change (FC)	A description of levels of gene expression. E.g., if a signal is twice as high in a breast tumour sample as in normal breast tissue, then the expression ratio of that gene is 2.

1.4.2.5 Post-microarray validation of microarray data

Whilst microarray gene expression analysis is a very powerful technology, it is fallible, and observed gene expression changes must be backed up by other methods, for example RT-PCR or Northern Blotting, and Western Blotting (Chuaqui et al., 2002). Also, knowledge of gene expression patterns is of limited use in most situations, unless backed up by functional studies which attempt to assign a function to the gene and place it in its biological context. To assist with the assignment of gene function, and to enable researchers to understand the broad gene expression in the cell, Gene Ontology (GO) is used. GO projects, such as that managed by the Gene Ontology Consortium (http://www.geneontology.org) attempt to "provide a controlled vocabulary to describe gene and gene product attributes in any organism" (Ashburner et al., 2000). The use of GO could, ideally, distinguish genes significantly related to the phenotypes being investigated, and secondly, identify groups of genes playing a coordinate role. So, for example, if 20% of genes from a candidate gene list were all annotated with the GO term "negative regulator of apoptosis" one could postulate that this process must be related to the phenotype studied.

1.4.3 Gene expression signatures for the prediction of clinical outcomes

1.4.3.1 The Amsterdam signature

The first gene signature predictive of solid tumour prognosis was proposed by Van't Veer et al. (2002), who investigated the gene expression of primary breast carcinomas from 98 patients. Of these patients, 34 developed metastases within five years and 44 remained disease-free after five years. They performed cDNA microarray analysis where RNA from each tumour was compared against pooled tumour RNA. Genes were considered differentially expressed if the fold-change was greater than 2 ($P < 0.01$) in more than 5 of the 98 tumours (i.e. in 5.1% of the samples). Cluster analysis, the grouping of genes according to their gene expression pattern, divided the tumours into two clusters. The clinical history of these tumours was then examined. In one cluster, the tumours were found to be predominantly from patients with a good prognosis (disease-free after five years). The tumours in the second cluster were predominantly from patients with a bad prognosis (metastases within five years). The clustering was also associated with the ER-alpha receptor status (used as an indicator of prognosis, where negative ER receptor staining is associated with bad prognosis). After the application of further clustering algorithms, they developed a gene expression signature containing 70 genes, believed to be indicative of poor prognosis.

In a subsequent study by the same group, the signature was tested on another cohort of breast cancer patients (van de Vijver et al., 2002). Here, 295 tumours were analysed by microarray and gene expression compared to the poor prognosis signature. They reported a strong correlation between tumours exhibiting the "good prognosis" signature, and survival and the absence of metastases. It outperformed all other methods currently used to predict clinical outcomes, at predicting the likelihood of distant metastases within five years, and thus, would improve the selection of patients for adjuvant systemic therapy. The Amsterdam signature was the first demonstration that gene expression profiling could be used to guide treatment decisions, and potentially result in fewer patients being under- or over-treated.

1.4.3.2 Other predictive gene signatures

Various other groups have derived predictive gene signatures, just two more of which are discussed here. Pawitan et al. (2005) used a similar approach to Van't Veer et al. to obtain a subset of genes (the Swedish signature) that could distinguish between good

and bad prognosis. They extracted RNA from 150 breast tumours (38 of which were from patients with a poor prognosis, 121 from patients with a good prognosis) and compared gene expression using Affymetrix oligonucleotide arrays. Only three out of the 64 genes in their dataset also appear in the Amsterdam signature. A third signature, the Rotterdam signature, was developed from a similar study on primary breast cancers by Wang et al. (2005). Likewise, there is only a three-gene overlap between the Amsterdam and Rotterdam signatures, and the three genes are not the same for all three signatures. A possible reason for the lack of cross-over was that different microarray technologies were used; cDNA microarrays in the development of the Amsterdam signature and oligonucleotide arrays in the development of the Rotterdam and Swedish signatures. Moreover, different genes were represented on each type of microarray. This indicates that conclusions drawn by microarray analysis are heavily influenced by the technology used. Secondly, the genetic variability of the cohorts of patients probably made it difficult to establish a common signature.

1.4.3.3 Current clinical value of predictive gene lists

There is a curious lack of overlap between the gene expression signatures of metastasis that have been described so far. Furthermore, the Amsterdam and Swedish signatures were unable to correctly predict outcomes when tested on each other's patient cohorts (Ein-Dor et al., 2006), and a more comprehensive re-analysis of seven major predictive gene lists by Michiels et al. (2005) was unable to produce gene lists that were any better at classifying patients than chance. Ein-Dor et al. (2006) performed mathematical modelling to investigate the extent to which patient genetic variability contributes to the data. They concluded that the problem underlying the unreliability of the recent reports is that there are not enough patients, and propose that thousands, rather than hundreds, of patient samples are needed to generate a reliable, robust predictive gene list.

To summarise, a significant body of work demonstrates the use of microarray analysis in the search for gene expression signatures that can be used to predict clinical outcomes and highlight potential drug targets. However, more work is needed in order to develop a gene expression signature with an acceptably low rate of misclassification of metastatic risk. Ideas emerging from the recent reports are that there is infrequent overlap between predictive gene lists developed by different groups, and that patient-based microarray analyses theoretically require thousands of well-curated samples to yield clinically useful gene expression signatures.

1.4.4 Gene expression analysis of animal models of metastasis

Animal models of metastasis can be used in conjunction with microarray analysis to search for genes that support the metastatic phenotype. The main advantages of animal models of metastasis are that they are tractable, and that the effects of certain genes can be tested by an *in vivo* metastasis assay, or by *in vitro* assays designed to test the ability of cells to perform metastasis-related behaviours such as migration or adhesion. The use of animal models can also avoid some of the challenges posed by patient-based analysis of metastasis. For example, the variability arising from using cohorts of patients with different genetic background can be avoided by using the same animal strain or, ideally inbred rats to reduce genetic variation in the host. Furthermore, in animals, it is possible to investigate the expression changes which mediate the progression from a benign to a metastasising tumour, by comparing primary and secondary tumours, whereas in patients, usually only one sample per patient is available.

1.4.4.1 Rho C as a metastasis promoter in melanoma cells

The first major report of gene expression analysis of an animal model of metastasis came from Clark et al (2000), who used *in vitro* selection to obtain highly metastatic variants of a poorly metastatic tumour cell line. They performed microarray analysis of the parental tumours and metastases, and found a number of genes with enhanced expression in metastases. Some of these genes, such as the cytoskeletal proteins Thymosin β4, Rho C, α-Catenin, α-Actinin and IQGAP, had potential roles in aspects of metastasis. The investigators tested Rho C by overexpression, or expression of a dominant negative Rho C, in the cells. The cells were subjected to a metastasis assay, where cells were injected into the tail vein of a mouse; this tested the stages of the metastatic cascade from the survival of the cells in the blood stream followed by extravasation and proliferation. The metastases were counted after some weeks. They found that Rho C overexpression increased the number of metastases and that expression of dominant-negative Rho C markedly decreased the number of metastases, indicating that Rho C activity contributes to the development of metastases in this model. This interesting finding demonstrates the value of microarray analysis in metastasis models.

1.4.4.2 Ezrin as a metastasis promoter in osteosarcoma

Khanna et al (2001) used a mouse osteosarcoma model to investigate genes involved in metastasis. The model was comprised of two clonally-related cell lines, with different

metastatic potentials. They found that 53 genes were significantly differentially expressed between the two cell lines. One of these, the ERM protein ezrin, was upregulated in the more aggressive cell line. They investigated the role of ezrin in metastasis (Khanna et al., 2004) by suppressing its expression in the more aggressive osteosarcoma cells and subjecting them to an *in vivo* metastasis assay, performed by intravenous tail injection of the cells. They found that the reduction in ezrin significantly reduced the development of pulmonary metastases.

1.4.4.3 Metastasis assays involving subcutaneous injection

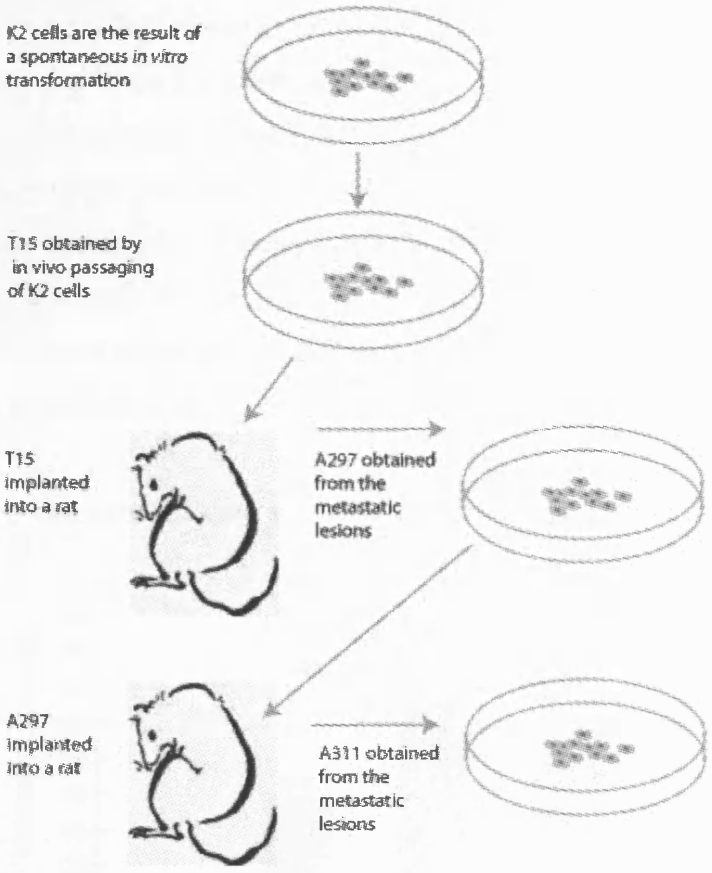
Other studies describe the gene expression of metastasis using a model where the cells were delivered to the animal by subcutaneous implantation. This approach, unlike intravenous injection, tests all parts of the metastatic cascade, including intravasation. Margalit et al. (2003) investigated the gene expression in pulmonary metastases that developed after subcutaneous injection with D122 lung carcinoma cells or B16-F10.9 melanoma cells into inbred mice. They found WISP-1 overexpression in the pulmonary metastases of both the cell lines, suggesting that WISP-1 overexpression may be important for metastasis. Li et al. (2003) also used the subcutaneous implantation approach, to develop highly metastatic hepatocellular carcinoma (HCC) cell lines by serial tumour progression in nude mice. The cell lines shed metastases throughout the animal at rates of between 70 and 100%. They performed microarray analysis of the newly developed and highly metastatic cell line in comparison to the parental cell line and found 25 differentially expressed genes. These genes were thought to potentially confer the ability to shed metastases. Another model using subcutaneous implantation into the mammary gland was used by Yang et al. (2004) to identify the transcription factor Twist as being a regulator of EMT, which is a common feature of metastatic cells. Finally, Wang et al. (2002b) suggested that the mRNA transporter ZBP-1 is important to the localisation of β-actin and hence motility, from gene expression profiling of a similar model.

1.4.5 Investigation of metastasis using a rat sarcoma model

I decided to investigate the gene expression of metastasis using a rat sarcoma model developed by Vesely et al. (Vesely, 1989; Vesely et al., 1987; Vesely and Weiss, 1973). The model was developed in inbred LEWIS rats and is comprised of four related cell populations (see Figure 1.13). K2 cells were derived by spontaneous transformation of embryonic fibroblasts *in vitro* (Vesely and Weiss, 1973). T15 cells were obtained from K2 cells by *in vivo* passaging. T15 cells were implanted subcutaneously into a rat and metastasis was allowed to occur; A297 cells were recovered from the metastatic lesions. Similarly, A297 cells were implanted subcutaneously into a rat and A311 cells were recovered from a resulting metastasis.

Similar models have proved useful for studies of metastasis in the past (Fidler, 1973) and are valuable because they are testable, have a clear history, and originate from the same parent tumour (thus enabling assessment of the features that caused cancer cells to progress through the metastatic cascade with different efficiencies). Further advantages of this model are that it is derived from inbred rats, and therefore the genetic background is similar and differences observed between the cells are likely due to differences in metastatic potential rather than differences between subjects. Also, our approach includes the effects of the immune system, which is known to play a role in metastasis, for example, tumour associated macrophages can contribute to ECM degradation. Furthermore, our chosen model involves subcutaneous implantation rather than tail vein injection of the tumours cells during the metastasis assay. This approach tests all of the steps of the metastatic cascade.

Figure 1.13: Development of the rat sarcoma model.

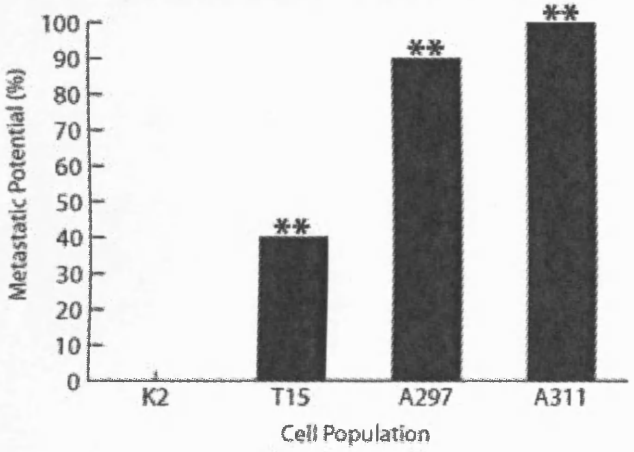


The rat sarcoma model of metastasis is the result of a combination of in vitro and in vivo passaging of K2 cells, which are spontaneously transformed rat embryonic fibroblasts. T15 are a subpopulation of a K2 culture, and A297 and A311 were obtained from the metastatic lesions arising after implantation of T15 and A297 cells.

The cancer in this model is sarcoma, which is a class of tumour arising from mesenchymal tissue, for example, fibroblasts, muscle and bone. Tumours arising from epithelial tissues, such as skin and mammary gland, are called carcinomas. Whilst sarcomas are less prevalent than carcinomas, sarcoma models of metastasis can offer insights into both sarcomas and carcinomas, because of the tendencies of epithelial cancers to undergo EMT (see Section 1.1.3.4).

All four rat sarcoma cell populations of the model are oncogenic, but have different metastatic potentials (defined as the percentage of primary tumours that shed metastases). To assess the metastatic potential of the four cell populations, 1 million cells were injected subcutaneously into inbred Lewis rats (Pokorna et al., 1994). In every case, primary tumours developed. After 4-6 weeks, the rats were sacrificed and inspected for metastases. The metastases localised primarily to the lung but were observed elsewhere. The metastatic potentials of the cells are shown in Figure 1.14. The K2 cells have no metastatic potential. However the T15, A297 and A311 cells have metastatic potentials of 40, 90 and 100% respectively.

Figure 1.14: Rat sarcoma cells with differences in metastatic potential.



*The cell populations K2, T15, A297 and A311 were subcutaneously implanted, in doses of 1 million cells, into inbred LEWIS rats. The metastatic potentials (the percentage of subjects which developed secondary tumours) are significantly different between the K2 cells and the T15, A297 and A311. The K2 cells have no metastatic potential, whereas the other three cell populations have metastatic potentials of 40, 90 and 100% respectively; **P < 0.01. The number of rats tested with the K2, T15, A297 and A311 cells were 20, 24, 10 and 10 respectively. This experiment was performed by Eva Pokorná (Institute of Molecular Genetics, Academy of Sciences of the Czech Republic).*

The following chapters describe the characterisation of the rat sarcoma cells, and the analysis of gene expression in the model. Following this, candidate genes indicated by the microarray were systematically tested for their role in behaviours which relate to metastasis.

Chapter Two

Materials and methods

Unless otherwise stated, all reagents were supplied by Sigma-Aldrich.

2.1 Cell culture

2.1.1 Sarcoma cells

Rat sarcoma cell populations K2, T15, A297 and A311 were a gift from Pavel Vesely (Institute of Molecular Genetics, Academy of Sciences of the Czech Republic). K2, full name LW13K2, are spontaneous neoplastic transformants from inbred LEWIS rats (Vesely and Weiss, 1973). T15, full name RPSL4T15, were developed from K2 by neoplastic progression *in vivo* and *in vitro* (Vesely et al., 1987). A297, full name A297Nb, were developed from T15 cells by *in vivo* selection, where cells were collected from a tumour in the rat with the highest number of metastases. A311, full name A337/311RP, were obtained from a retroperitoneal metastasis that arose when the A297 cells were applied to the rats. The cells were maintained in MEM with Hanks' salts supplemented with 10% bovine serum (SML, Germany), 0.09% sodium bicarbonate, 1 mM glutamine and phenol red indicator, at 37°C with 5% CO_2 in a humidified incubator, and passaged every 2-3 days. Passaging of cells involved a brief wash in 0.02% trypsin/5 mM EDTA followed by incubation in 0.02% trypsin/5 mM EDTA for 0.5-1 minute to detach cells. Further trypsin/EDTA activity was inhibited by the addition of fresh culture medium to the suspension. The cell suspension was used to prepare new cultures. For motility and chemotaxis assays, cells were seeded onto coverslips at a density of ~3500 cells per cm^2. For transfections, cells were seeded at a density of ~7000 cells per cm^2. For routine subculture, cells were seeded at a density of ~4000 cells per cm^2.

2.1.2 HeLa cells

HeLa cells were cultured in E4 medium supplemented with 10% foetal bovine serum and phenol red indicator, and maintained at 37°C with 10% CO_2 in a humidified incubator. Cells were passaged every 2-3 days as described for the sarcoma cells, and seeded for transfections at ~25,000 per cm^2 and for routine subculture at ~16,000 cells per cm^2

2.2 *Immunocytochemistry and confocal microscopy*

2.2.1 *Immunocytochemistry*

Cells growing on 13 mm diameter, 170 μm thick glass coverslips were fixed in 4% FA/PBS for 10 minutes, washed and permeabilised by incubation in cold 0.1% Triton/PBS for 10 minutes. Cells were incubated with 1% bovine serum/PBS to block non-specific sites prior to a 45 minute incubation with mouse α-Actinin antibody (Santa Cruz) or mouse CASK antibody (Chemicon International) diluted 1:400 in PBS. Cells were then washed three times with PBS over the course of 10 minutes, then incubated with Alexa 488-conjugated anti-mouse IgG diluted 1:400 in PBS, and rhodamine phalloidin diluted 1:2000 (both from Molecular Probes) for 30 minutes, and washed three times with PBS for 10 minutes. Coverslips were mounted by inversion onto 8 μl Mowiol on a microscope slide, and allowed to dry overnight and then stored at 4°C. Mountant was prepared by shaking 12 ml Citifluor AF1 (Citifluor Ltd.) and 2.4 g Mowiol 4-88 (Calbiochem) for 2 hours. Tris-HCl was then added to a final concentration of 100 mM and incubated at 50°C with mixing until dissolved. The solution was then filtered through a 0.45 μM filter and stored at -80°C.

2.2.2 *Confocal microscopy of fixed cells*

Confocal microscopy was performed using a Zeiss LSM 510 microscope using a 63×, NA1.4, phase-contrast objective lens. Z-stacks were acquired using Zeiss software, taking an optical section every 0.2 μm from the adhesion plane to the top of the cell. Images were optimised by averaging 8 scans. The final stack size was approximately 146.2 μm × 146.2 μm × ~2 μm; 1024 × 1024 × 11 pixels.

2.2.3 *Analysis of actin arrangements in phalloidin-stained cells*

2.2.3.1 *Acquisition*

Images of the sarcoma cells were randomly selected. An area of the coverslip approximately 2 mm × 2 mm was scanned at low magnification and the images placed together in a tile scan, which gave a low magnification image of the region. Groups of cells were visible but at too low a magnification to be able to tell whether or not stress fibres were present. This eliminated biased acquisition. Groups of cells in the tile scan were marked and fields acquired at × 25 magnification.

Images of the HeLa cells were acquired in a semi-random way where groups of cells were found by looking at the wide-field phase contrast image, from which it was not possible to tell whether or not stress fibres were present.

2.2.3.2 Image analysis

Images of phalloidin-stained cells were subjected to a double-blind analysis of F-actin structures using Metamorph® software (Molecular Devices Corporation). Randomly-named images were opened in Metamorph, and each cell was inspected and classified to an F-actin category by manually selecting the cell with the mouse. All Z-sections of the image, from the adhesion plane to the top of the cell, were inspected for actin structures. HeLa cells were categorised as containing either "stress fibres", where at least one stress fibre was observed, or "disordered actin", where filaments but no stress fibres were observed, or "cortical actin", where actin was restricted solely to the cell cortex. Sarcoma cells were categorized as containing either "stress fibres" or "no stress fibres". Using a dynamic data exchange all cell assignments made were directed logged into an Excel spreadsheet (Microsoft). The correct descriptions of the images were then revealed and the percentage of control and treated cells falling into each actin category was calculated.

2.3 Cell motility assays

2.3.1 Dunn chemotaxis assay

Chemotaxis was assessed using the Dunn chemotaxis chamber (Zicha et al., 1991). The assembly of the Dunn chemotaxis chamber is summarised in Figure 2.1, and described in detail here. 18 mm × 18 mm, 170 μm thick glass coverslips were prepared for use in the Dunn chemotaxis chamber by immersion in a mixture of 60% HCl and 40% ethanol for at least 20 minutes, followed by immersion in distilled water for 20 minutes, and a final immersion in 70% ethanol for sterilisation. Coverslips were dried with compressed air filtered through a 0.45 μM filter (Millipore). Cells were seeded on coverslips three days before the experiment, and grown to 5-10% confluence. They were then serum-starved by incubation for five hours in medium containing 0.5% BS. A sterile Dunn chamber was placed on a flat clean surface and 200 μl phenol red-free starvation medium was placed into the wells of the chamber. The coverslip of cells was picked up with fine sterile forceps and inverted gently onto the pool of medium flooding the wells. Using a tissue, the coverslip was positioned so that there was a gap of approximately 1 mm on one side of the coverslip allowing access to the outer well. The

other three sides of the coverslip were dried by applying gentle pressure with a tissue, and sealed with a molten wax mixture. The wax was comprised of equal volumes of bleached white beeswax (Acros Organics), paraffin wax (Fisher Scientific) and soft yellow paraffin (Fisher Scientific), mixed and heated to 60°C. The outer well was drained of starvation medium by drawing the liquid out of the chamber with a tissue. The central well of the chamber still contained starvation medium. The outer well was then washed once with fresh medium to remove debris or factors released by cells crushed during the assembly of the chamber. Then starvation medium supplemented with 60 ng/ml human PDGF-β (Peprotech EC) and 80 ng/ml human IGF-I (Peprotech EC) was added to the outer well. The open side of the chamber was dried and sealed with the wax mixture. The assembled Dunn chambers were rinsed gently with distilled water and dried with compressed air to remove debris and crystals from the coverslip.

Figure 2.1: Assembly of the Dunn chemotaxis chamber.

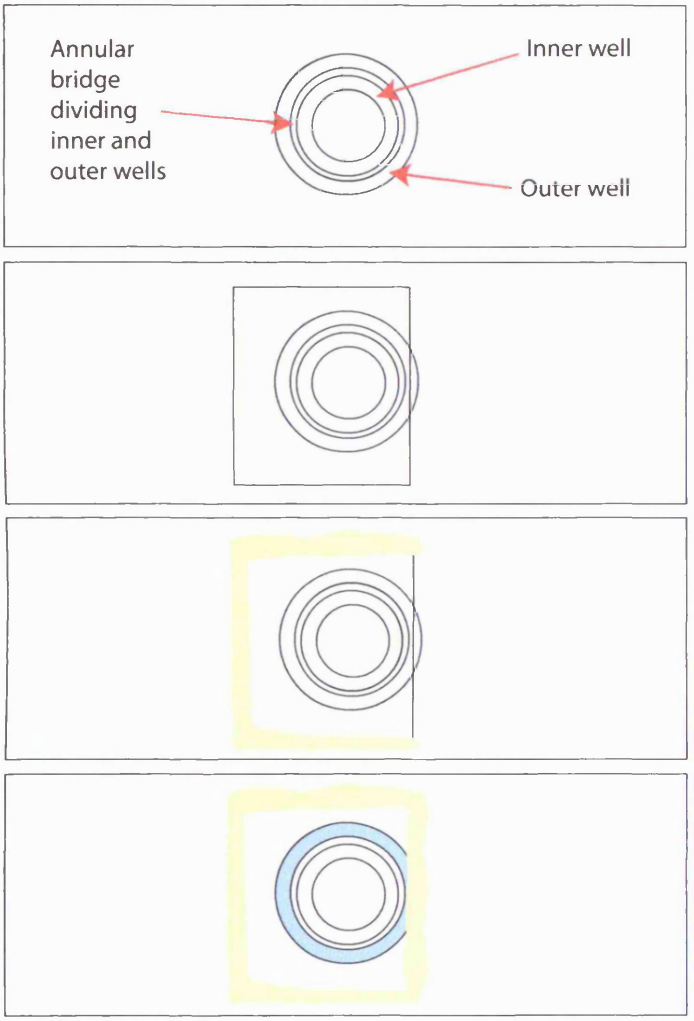
The Dunn chemotaxis chamber is a modified bacteriological counting chamber with two concentric cells separated by an annular bridge, the top of which is 0.02 μm lower than the rest of the slide.

(A) The wells are flooded with approximately 200 μl starvation medium

(B) A coverglass with cells is inverted onto the medium, covering both wells, leaving a filling gap of about 1 mm on one side.

(C) The other three sides of the chamber are dried, and sealed with wax.

(D) The outer well is drained and filled with medium containing chemoattractant. Finally the open side of the chamber is dried, and sealed with wax.



(A) The wells are flooded with approximately 200 μl starvation medium and (B) a coverslip with cells is inverted onto the medium and positioned so that it covers both wells, leaving a filling gap of about 1 mm on one side. (C) The three edges of the coverglass without the filling gap are dried and sealed with wax. (D) The medium is removed from the outer well; the starvation medium remains in the inner well since it has been sealed in by the coverglass. The outer well is then rinsed with starvation medium to remove any cell debris, and refilled with medium containing chemoattractant. Finally the open side of the chamber is dried and sealed with wax.

Within 30 minutes a concentration gradient develops at the bridge region, due to the presence of growth factor-supplemented medium on one side of the bridge and starvation medium on the other. Cells overlying the bridge region are exposed to this concentration gradient and will migrate the other well, if they are chemotactic to the growth factors therein. All media for these experiments were used at 37°C, and equilibrated to 5% CO_2 in the incubator.

2.3.2 Random walk assay

When investigating general cell motility rather than chemotaxis, the random walk chamber (Figure 2.2) was used. The random walk chamber is composed of a microscope slide out of which a 10 mm diameter circle has been cut. A coverslip (thickness ~230 μm) was glued to the slide. This results in a slide containing a 10 mm diameter well that can be filled with medium. Coverslips of cells can be inverted over the well and fixed into position using wax to create a hermetically sealed environment. Cells were grown to approximately 5-10% confluence on square 18 mm, 170 μm thick coverslips (prepared as described for the Dunn chamber) for three days. They were starved by five hour incubation in medium containing 0.5% BS. A sterile random walk chamber was placed on a flat clean surface and 200 μl phenol red-free medium, containing 10% BS, was placed into the well. The coverslip of cells was picked up with sterile fine forceps and inverted gently onto the well. Each side of the coverslip was dried by applying gentle pressure with a tissue, and then sealed with the molten wax mixture described before. Random walk chambers were rinsed gently with distilled water and dried with compressed air to remove debris and crystals from the coverslip.

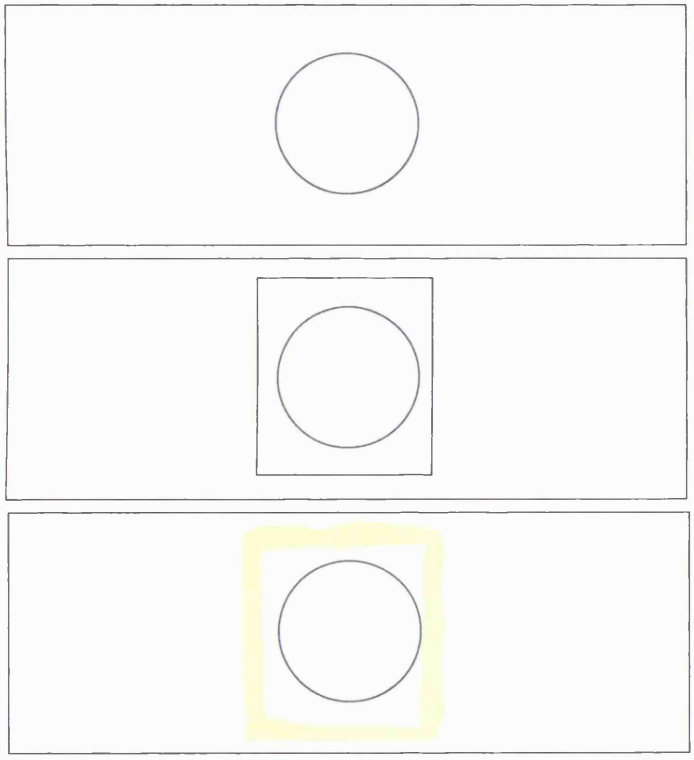
Figure 2.2: Assembly of the random walk chamber.

The random walk chamber is comprised of a microscope slide with a 10 mm diameter hole drilled in the centre. A coverglass is glued to the slide to create a 10 mm diameter well.

(A) The well is flooded with approximately 200 µl medium.

(B) A coverglass of cells is inverted onto the well.

(C) The edges of the coverslip are dried, and sealed with wax.



(A) The well is flooded with approximately 200 µl medium (warmed and gassed to 37°C, 5% CO2) and (B) a coverglass with cells is inverted over the well. (C) The edges of the coverglass are dried with a tissue and sealed with wax.

2.4 Low light level digital microscopy

2.4.1 Equipment and acquisition

Movies were acquired with a 10× NA 0.3 phase contrast objective on an Axiovert 135TV inverted microscope (Zeiss) or a 10× NA 0.17 phase contrast objective on a Diaphot or an E1000 upright microscope (Nikon). The microscopes were equipped with sensitive CCD cameras (Orca ER, Hamamatsu Photonics UK Limited). A motorized stage (Märzhäuser Wetzlar GmbH & Co) and piezo focus (Physic Instrumente) on the E1000 microscope, and a motorized stage (Prior Scientific Instruments Ltd.) and piezo focus (Physic Instrumente) on one of the 135TV inverted microscopes, allowed multiple fields to be acquired. The microscopes were housed in Perspex chambers developed to accurately maintain a temperature of 37°C. The Dunn or random walk chamber was placed on the microscope stage and suitable fields found. For motility studies, a suitable field contained relatively sparsely seeded cells, distant from any cells crushed during the assembly of the chamber. AQM software (Kinetic Imaging) was used to collect images every 5 minutes, of either phase contrast only, or sequential phase contrast and GFP/Cy3 images. A total of 200 time-points were collected resulting in films that were 16.6 hours in duration.

2.4.2 Evaluation of films of migrating cells

Figure 2.3 summarises the method used to assess cell motility. Films were reviewed in Motion Analysis software (Kinetic Imaging) and cell trajectories were produced by interactive tracking of cell translocations. The speed and chemotaxis of the cells were evaluated using an algorithm (Zicha and Dunn, 1995) implemented in Mathematica® (Wolfram Research, Inc.). In this algorithm, cell trajectories were aligned to a common origin. For each film, the direction of the increasing concentration of the growth factors in the gradient was determined according to the position of the outer well in the images. Trajectories from individual films were normalized by rotation using the direction angle. In multiple field acquisition it was impossible to position the outer well at the top of the computer screen for each field. After rotation, the angle of the direction of cell movement was calculated. Calculation of the angle depended upon the cell migrating a distance of at least 70 μm. The distributions of the tracks were represented by circular histograms, where the red arrow indicates the mean direction of the cell migration, and the green wedge indicates its 95% confidence interval. The statistical comparison between chemotactic responses of the cell populations was based on ANOVA applied to normalised displacement in the direction of the gradient. Occasionally, due to slight differences in cell culture conditions, the speed of the control cells varied from one culture to another. In such cases, the speeds were normalized using the average speed of uninjected cells from the same field. The Mathematica® Notebook used for the processing of cell tracks can be found in Appendix 1.

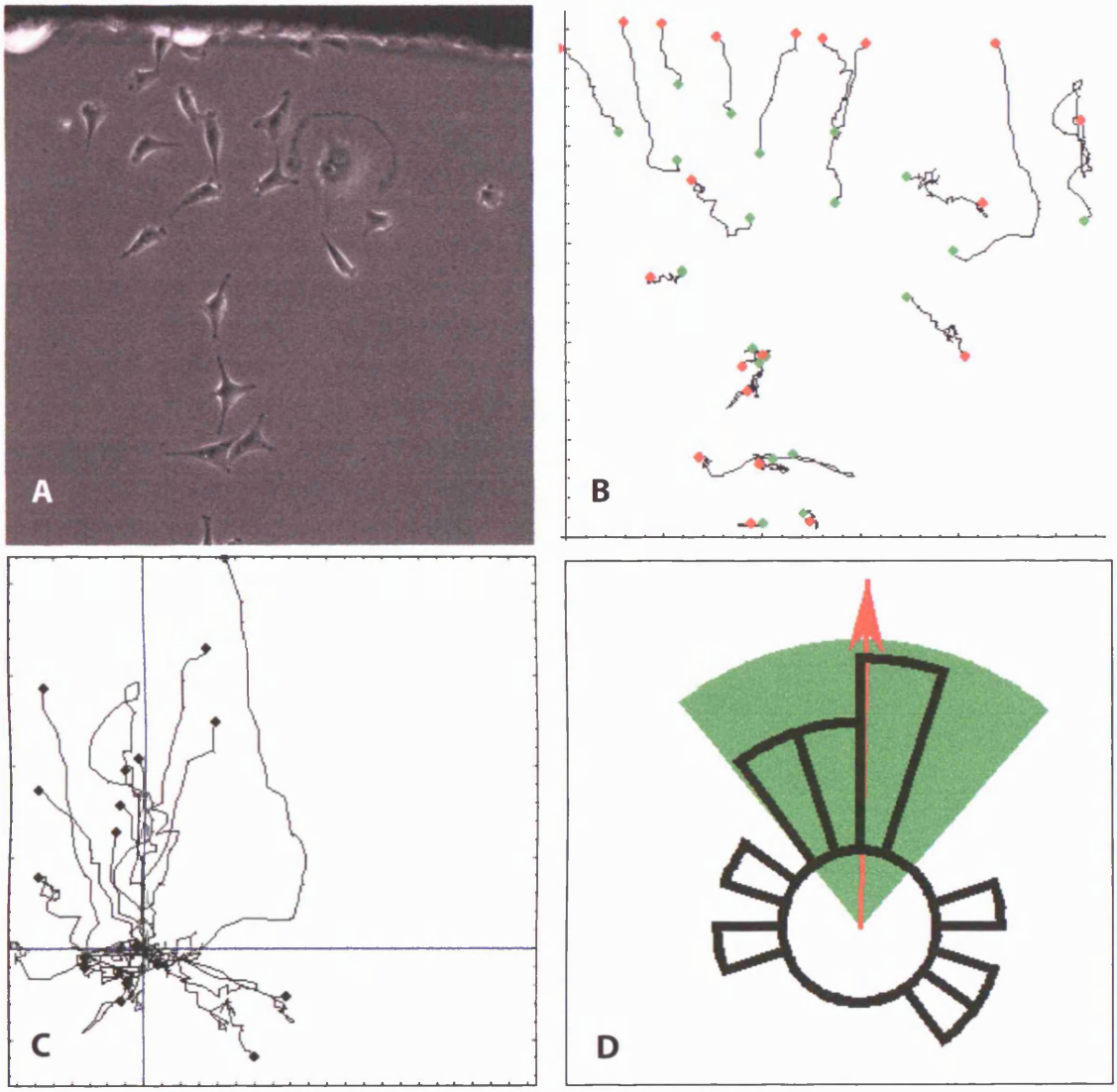


Figure 2.3: The evaluation of cell motility.

To assess cell motility, cells were filmed overnight on a low light level digital microscope. Cell translocations were tracked using Motion Analysis software (Kinetic Imaging), and their trajectories analysed using an algorithm developed in Mathematica to evaluate cell speed and the distribution of the directions of cell trajectories.

(A) A field of A297 cells in the Dunn Chamber. Here, the outer well (containing growth factors) is visible at the top of the field and the inner well (containing starvation medium) is at the bottom.
(B) Cell trajectories, where the green dot represents the position of the cell at the first frame of the film and the red dot represents its position at the end of the film.
(C) Data are imported into Mathematica and the origins of the trajectories are aligned.
(D) The data are processed in Mathematica and the distributions represented by circular histograms, where the red arrow indicates the mean direction of cell migration, and the green wedge indicates its 95% confidence interval.

2.5 Gene expression analysis

2.5.1 Overview of Affymetrix GeneChip® system

Microarrays measure the expression of thousands of genes in parallel. In this study, the microarray experiments were performed using Affymetrix Rat 230A GeneChips®. The GeneChip® contains an array of oligonucleotide probes bound to the glass of the chip. Each probe is designed to hybridise to a transcribed sequence of the rat genome. The Rat 230A GeneChip® is about 1.3 cm^2 and contains approximately 300,000 oligonucleotide probes, which together measure the expression of approximately 15,000 transcribed sequences. Briefly, RNA is extracted from the sample, fragmented, labelled with biotin and allowed to hybridise to the microarray overnight. This RNA hybridises to its complementary probes sequences on the GeneChip® array. The biotin reacts with streptavidin-phycoerythrin, use to stain the arrays, and the amount of fluorescence, which corresponds to the amount of hybridisation, is recorded by a camera. The extent of hybridisation is an indication of the amount of transcript in the sample.

2.5.2 Experimental design

The first microarray experiment was two dimensional. I analysed (1) gene expression between a metastatic and a non-metastatic sarcoma cell population and (2) their gene expression responses to the chemotactic growth factors PDGF and IGF. The project was then extended to include two more cell populations, plus the primary tumours that developed when the cells were subcutaneously implanted into rats. Three biological replicates were performed for each condition throughout.

2.5.3 Sample collection

2.5.3.1 Cultured cells

Sub-confluent, early passage cells were harvested by detachment with trypsin after being made quiescent in medium containing reduced (0.5%) serum for five hours, followed by 1) no further treatment, 2) stimulation for 30 minutes with 30 ng/ml PDGF and 40 ng/ml IGF, or 3) stimulation for 180 minutes with 30 ng/ml PDGF and 40 ng/ml IGF. Samples were checked for mycoplasma by agar colony growth, and found to be free of contamination.

2.5.3.2 Tumours

One million cells were injected subcutaneously into rats and primary tumoursbegan to develop. The animals were sacrificed 4-6 weeks later, and 1 cm^3 of primary tumour was excised and stored at -20°C in RNAlater (Qiagen).

2.5.4 Preparation of RNase-free reagents

2.5.4.1 DEPC water

Diethylpyrocarbonate (DEPC) treatment inactivates RNases. DEPC was added to distilled water at 0.1%. The bottles were shaken thoroughly, incubated overnight at room temperature, and then autoclaved.

2.5.4.2 Preparation of RNase-free reagents and the RNase-free environment

All reagents, for example buffers, were treated with DEPC after being prepared, or made up with DEPC-treated water. Equipment, instruments and surfaces used in the preparation of RNA were cleaned with 0.3% hydrogen peroxide or RNase Zap Wipes (Ambion), followed by rinsing in DEPC-treated water.

2.5.5 RNA extraction

The cultured cells were homogenised in Qiashredder columns. Tumour samples were tough and accordingly received a different homogenization; after being removed from the RNA stabilization solution and weighed, they were placed in 750 µl RLT lysis buffer (provided with the Qiagen kit) per 100 mg of tissue. They were homogenized for 40 seconds with a rotor-stator homogeniser, sonicated for 20 seconds, and the resulting lysates then cleared of cell debris by centrifugation. Total RNA was extracted with the RNeasy Mini-Kit (Qiagen) according to the manufacturer's instructions. The concentration and purity of the RNA was determined using readings taken at 260 nm and 280 nm on a Nanodrop spectrophotometer.

2.5.6 RNA agarose gel electrophoresis

Before the RNA was further manipulated, its quality was verified by visualization of 28S and 18S ribosomal RNA bands by gel electrophoresis. 1 µg RNA was loaded on a 1% TBE agarose gel supplemented with 1 µg/ml ethidium bromide. The gel was run at 60 volts for one hour, and the DNA visualized using a UV transilluminator. When clear, undegraded bands of 28S and 18S ribosomal RNA were seen at an intensity ratio of approximately 2:1, the RNA was considered to be of high quality. Smeared bands

indicated that the RNA had been degraded at some stage, and was not acceptable for use in further experiments.

2.5.7 Preparation of biotin-labelled, fragmented cRNA

The extracted RNA was reverse transcribed to obtain cDNA. This was then subjected to *in vitro* transcription to obtain cRNA, which was labelled with biotin and fragmented ready for hybridisation to the Affymetrix GeneChip® array.

2.5.7.1 First strand synthesis

10 µg of total RNA and 5 µM T7-Oligo$(dT)_{24}$ primer (Affymetrix) were denatured by incubation at 65°C for 10 minutes, and then allowed to anneal by cooling on ice for 5 minutes. The RNA was then reverse-transcribed into cDNA using Superscript II Reverse Transcriptase (Invitrogen) and incubated at 42°C for an hour.

2.5.7.2 Second strand synthesis

The first-strand reaction was incubated at 16°C for 2 hours with 1 mM dNTPs, Second Strand Buffer, 10 units *E.coli* DNA Ligase, 40 units DNA Polymerase I and 2 units *E.coli* RNase H (all supplied by Invitrogen). Following this, 10 units T4 DNA polymerase were added, and after a further 5 minutes at 16°C, the reaction was terminated by the addition of 10 µl 0.5M EDTA. The cDNA was cleaned-up using the GeneChip® Sample Cleanup Module (Affymetrix) according to the manufacturer's instructions.

2.5.7.3 Synthesis of fragmented cRNA

In vitro transcription (IVT) was performed using the Enzo BioArray High Yield Transcript Labelling Kit (Enzo Diagnostics) according to the manufacturer's instructions. The cDNA was incubated at 37°C for 5 hours with feaction buffer, biotin-labelled ribonucleotides, DTT, RNase inhibitor mix and T7 RNA Polymerase, resulting in an approximately 100-fold amplification of RNA. Unincorporated dNTP were removed with the GeneChip® Sample Cleanup Module (Affymetrix) according to the manufacturer's instructions. The concentration and purity of the cRNA was determined using readings taken at 260nm and 280nm on a Nanodrop Spectrophotometer. 25 µg cRNA was incubated in fragmentation buffer (Affymetrix) at 94°C for 35 minutes.

2.5.7.4 Confirmation of sample quality

The purity of the cRNA was determined by the ratio of absorbance at 260/280 nm. Samples with 260/280 ratios between 1.9 and 2.1 were acceptable. A 1.2%

formaldehyde agarose gel was run to check the quality of the starting total RNA, cRNA and fragmented cRNA of all the samples, alongside a 0.24-9.5 Kb RNA ladder (Invitrogen) and RNA Century Size Markers (Ambion).

2.5.8 Hybridisation and scanning

Hybridisation of cRNA targets to the array and scanning were carried out at the Paterson Institute in Manchester, UK. Briefly, the cRNA was processed using an Affymetrix GeneChip® Instrument System according to the manufacturer's instructions. Briefly, spike controls were added to 20 μg fragmented cRNA and then allowed to hybridize to the GeneChip® arrays overnight. Arrays were then washed and stained with streptavidin-phycoerythrin, and scanned on an Affymetrix GeneChip® scanner to generate an image of the entire GeneChip® array. Each spot (containing a probe set) on the GeneChip® array is 20 μm^2 and was scanned at a resolution of 3 μm, giving an average of 49 pixels per spot. After scanning, array images were assessed by eye to check that the scanner was aligned correctly and to check that there were no significant bubbles or scratches on the chip surface.

2.6 Analysis of microarray data

Data were analysed with Affymetrix Microarray Suite (MAS 5.0) and with an algorithm developed in Mathematica® (shown in Appendix 2).

2.6.1 Analysis of microarray images in MAS 5.0

Analysis of microarray images was carried out at the Paterson Institute. The raw scanned images were analysed with Affymetrix Microarray Suite (MAS 5.0) software. Individual features on the GeneChip® were identified and a grid was overlayed, to separate each spot from its neighbours. The average intensity of each spot was calculated using the pixels at the centre of each grid-square and not the pixels at the outer edge. The actual signal intensity was calculated from the difference in brightness between the PM and MM probes in each probe set (see Section 1.4.2.3 for a detailed explanation of the PM/MM probe design). Once signal intensity had been calculated, the transcript was assigned as absent, marginal or present.

2.6.1.1 Spike controls and housekeeping gene controls

The array included a set of control probes to facilitate normalization and scaling of the experiments. Control probe hybrisations were assessed kindly by Yvonne Hey at the Paterson Institute. 3'/5' ratios for GAPDH and beta-actin were found to be within

acceptable limits (from 0.97 to 1.72), and spike controls were present on all chips, at the expected intensities. When scaled to a target intensity of 100 (using Affymetrix MAS 5.0 array analysis software), scaling factors for all arrays were within acceptable limits (0.19-0.405), as were background, Q values and mean intensities.

2.6.2 Data analysis in Mathematica®

Signal intensity data were subjected to a number of processes leading to the production of lists of candidate genes. The algorithm is displayed in full in Appendix 2 and is described briefly here.

2.6.2.1 Exclusion of unexpressed genes from the data analysis

Signal intensity data were imported into a Mathematica® Notebook developed in the laboratory (Appendix 2). Transcripts which had been scored as absent by MAS 5.0 on all three replicates were excluded from the gene expression analysis.

2.6.2.2 Normalisation

Normalisation was needed to correct for systematic differences in intensity between arrays that did not represent true biological variation. We used the median intensity of each array to normalise the arrays to each other.

2.6.2.3 Calculation of expression ratios

Expression ratios (ER) were used to consider the difference in gene expression between two samples. For example, ER > 2 means that expression of a gene is twice as high in one sample compared to another. The mean expression value of the three replicates was used in the calculation of ER. In the first experiment, we determined the ER between the different cell populations, and between the responses of the populations to PDGF/IGF. In the extended microarray experiment, we also looked at the ER between the cell populations and their corresponding primary tumours.

2.6.2.4 Calculation of statistical significance

To determine the statistical significance of each ER, the *t* test was applied to the expression ratio data (not the raw intensity data) using the formula

$$t = \frac{\overline{X}_1 - \overline{X}_2}{s_{\overline{X}_1 - \overline{X}_2}}$$

where s is the standard deviation, the numerators are the sum of the means of each group and the denominator is the standard error of the difference between the two means. The resulting t value was used to find the P-value from a table of values.

Because of the large number of samples being tested, the P-value was corrected with the Benjamini-Hochberg algorithm (Benjamini and Hochberg, 1995) to avoid an unacceptably high number of false positive calls.

2.6.2.5 Generation of candidate gene list from the initial microarray experiment

Here we made two comparisons, (1) between the K2 and A297 cell populations and (2) between starved cells and cells treated with the growth factors. Different ER and significance criteria were applied to each dimension. Changes in gene expression between the populations were defined by ER > 1.6 ($P < 0.05$). Changes in gene expression in the growth factor response were defined by the less stringent parameters of ER > 1.3 ($P < 0.1$). Cluster analysis was performed to group genes according to shared expression patterns. The gene expression ratios were represented by a naturalistic colour scale indicating the size of the ER. In this colour scale, black represented the lowest values, and increasing ER values are in cyan followed by blue. ER = 1 (no change) are represented by green, and expression ratios higher than 1 (where the gene was overexpressed in the A297) were represented by yellow, orange and then red. The statistical significance is indicated as follows; *$P < 0.1$, **$P < 0.05$, *** $P < 0.01$. Larger expression ratios were not believed to necessarily indicate a more significant result; hence the inclusion of statistical significance when compiling the candidate gene lists. Genes were selected for further study on the basis of their expression pattern, statistical significance, and any information available in the literature.

2.6.2.6 Generation of candidate gene list from the extended microarray experiment

Expression ratios of the response to growth factors were calculated and the t-test applied. It became clear that there were no significant differences in gene expression in response to PDGF/IGF. Therefore, the three time-points were averaged and used to calculate the ER between the four cell populations, with ANOVA applied to calculate the statistical significance with the Benjamini-Hochberg algorithm. ER between metastatic and non-metastatic cells were then calculated by dividing the value for the T15, A297 and A311 populations by the value for the non-metastatic, K2 cell population. The criteria used for the final selection of candidate genes were ER > 2.5 ($P < 0.05$), with no significant difference in gene expression pattern between the cultured cells and the primary tumours. This generated a list of twenty-three genes. To produce a list of eighty differentially expressed genes, less stringent parameters, ER > 2 ($P < 0.05$) were applied to the data. To display the patterns of gene expression,

normalised absolute gene expression values were calculated, logged, and plotted to show the cell and tumour data for the same gene in one graph.

2.7 RT-PCR

2.7.1 Preparation of cDNA by reverse transcription

RNA was extracted from the cells of interest as described for the microarray. 10 μg of total RNA and 5 μM oligo(dT) primer (Affymetrix) were denatured by incubation at 65°C for 10 minutes, and then allowed to anneal by cooling on ice for 5 minutes. The RNA was then reverse-transcribed into cDNA using Superscript II Reverse Transcriptase (Invitrogen) and incubated at 42°C for 60 minutes. The cDNA was then treated with 2 units of RNase H (Ambion) for 20 minutes at 37°C.

2.7.2 PCR

Fragments of each gene of interest were amplified from the template cDNA by PCR using Taq polymerase (produced by Cancer Research UK) or Qiagen's Taq polymerase. 10% of the reverse-transcription product, 1 × Taq buffer, 0.2 mM dNTPs, 20 pmol forward primer, 20 pmol reverse primer, and 1 unit of polymerase were used in each reaction. PCR reactions were performed using a PTC-200 Peltier thermocycler (MJ Research) using the following protocol; an initial denaturation at 94°C for 2 min, followed by 30 cycles of 1 minute of denaturation at 94°C, 1 minute of annealing at 60°C, primer extension at 72°C for 1 minute per kilobase of product. A final extension step was performed at 72°C for 5 minutes to complete any unfinished polymerisation. PCR products were resolved on a 1% TBE agarose gel supplemented with ethidium bromide, and visualized by UV transillumination.

2.7.3 RT-PCR primers

2.7.3.1 Primers used for the validation of microarray data

The sequences of the primers used for RT-PCR for validation of microarray data are as follows, *alpha actinin-1* forward 5'- ACCAACTCAGCCACCTGC-3',
alpha actinin-1 reverse 5'- CTCGGCTTGGTCAGGTGG-3';
Bmp2 forward 5'- TCAAGCCAAACACAAACAGCC-3',
Bmp2 reverse 5'- ACGTCTGAACAATGGCATGA-3';
crmp4 forward 5'- GTGAGTGGAAGGGGTGTTGT-3',
crmp4 reverse 5'- GGTTGGGTCTTGGCATTTTA-3';

GAPDH forward 5'- TGCTGAGTATGTCGTGGAGTCT-3',
GAPDH reverse 5'- CCCTGTTGCTGTAGCCATATTC-3';
mmp2 forward 5'- CAAGCCCAAGTGGGACAAGAACC-3',
mmp2 reverse 5'- CAGGTCCCAATGTCGGTGTG-3';
mmp3 forward 5'- ATGAAAGGGCTCCCAGTCC -3',
mmp3 reverse 5'- AAACTCCAACTGTGAAGATCCGC -3';
mtap6 forward 5'- ATAAACTGGCTGAGGCGAAA-3',
mtap6 5'- ACTACAGGGGCTTGGTCCTT-3';
robo forward 5'- AGGGGAGAGAAGCTCTGGAC-3',
robo reverse 5'- TTGACGTTGGGAAAGTAGGC-3';
tpm forward 5'- AGGAAGCCTCACGAAAACAA-3',
tpm reverse 5'- TCCTCTCGCAGTGCTACTCA-3';
Adamts1 forward 5'-GATGGTTTACAGGCTGCCTTC-3',
Adamts1 reverse 5'-TTGTTTGGCACACCAGTAAGC-3';
Ankyrin 3 forward 5'- CCGACTCCCTCAGACACTACA-3',
Ankyrin reverse 5'- GTGTTCCTTCCAGGTCTCTCC-3';
Bk forward 5'- GACCTGATCGCCATAAGGAAG-3',
Bk reverse 5'-GGTTTTGAAGTGGGGAATCAA-3';
Cask forward 5'-ACCATTCGGAAAATCCATGAG-3',
Cask reverse 5'- CTGACACAAGGCCGATAACAA-3';
Cspg4 forward 5'- ACCATCCAGAGAGCCACAGTA-3',
Cspg4 reverse 5'- AGCAGGACGTTAGTGAGGACA-3';
Epb41l3 forward 5'- CATCCAGCAGCAAACTCTCAC-3',
Epb41l3 reverse 5'- GTCACGAAGGAACAGGGTAGG-3';
GAPDH forward 5'- TGCTGAGTATGTCGTGGAGTCT-3',
GAPDH reverse 5'- CCCTGTTGCTGTAGCCATATTC-3';
Grem1 forward 5'- GATGACTGAGAGCGTTGTTCG-3',
Grem1 reverse 5'- GACCCAGTCACCTTTCTCTGG-3';
Plk2 forward 5'- AACTTGGCCAATGCTCTGTTT-3',
Plk2 reverse 5'- AAGAGCATGTTCAGGGCGTAT-3';
Ril forward 5'- AGCAGGCCTGAGAACAAGAAC-3',
Ril reverse 5'- TAGCGGAAGGATCCAGACTGT-3';
Trax forward 5'- AACGCTTGCTATGCCCTTAAA-3',
Trax reverse 5'-CCTTCCACCCAAAATGTCACT-3'.

2.7.3.2 Primers used for the detection of epb41l3 isoforms

The sequences of the primers used for the detection of isoforms of *epb41l3* in the sarcoma cells are, forward 5'-ATGTCTCGCAGCTTGGATGGA-3', and reverse 5'-GCTTCTCCCCAGAAGATTGTTTAGTCT-3'.

2.8 *Immunoblotting*

Immunoblotting (Western blotting) was performed to verify protein expression.

2.8.1 *Preparation of cell lysates*

Subconfluent, early passage cells were detached by incubation with trypsin and centrifuged at 2000 rpm for 10 minutes. Supernatant was removed and the cell pellets were washed with PBS and lysed for 15 minutes in cold RIPA buffer (1x PBS, 1% Nonidet P-40, 0.5% sodium deoxycholate, 0.1% SDS) containing protease inhibitors (100 μg/ml PMSF, 50 KIU aprotinin and 1 mM sodium orthovanadate). Cell lysates were cleared by centrifugation at 10,000 g for 10 minutes. Total protein concentration was measured using the Coomassie Assay (Pierce) according to the manufacturer's instructions.

2.8.2 *SDS-Polyacrylamide gel electrophoresis and transfer*

Electrophoresis and transfer were carried out using an XCell SureLock apparatus (Invitrogen). Equal amounts of protein, typically 20 μg, were loaded onto a 4-12% Nupage Bis-Tris polyacrylamide mini-gel (Invitrogen) and electrophoresed at 125 V for approximately 90 minutes. Proteins were transferred onto a PVDF membrane (Immobilon P, Millipore) by the application of 200 V for at least 60 minutes. Membranes were stained briefly with Ponceau-S solution to verify that the protein transfer had been successful.

2.8.3 *Visualisation of proteins*

Membranes were blocked overnight in 5% non-fat dried milk (Marvel). They were then incubated with primary antibody for 1 hour. Antibodies to α-Actinin (Santa Cruz), CASK (Chemicon International) and 4.1B (kindly provided by Narla Mohandas, New York Blood Center) were used at a dilution of 1:2000. Antibodies to actin and GAPDH (both from Abcam) were used at a dilution of 1:10,000). Primary antibodies were washed three times in TBST over 30 minutes. Blots were then incubated with HRP-conjugated secondary antibody (all from Amersham Biosciences) diluted 1:5000 for 30 minutes followed by TBST washes as before. Protein bands were visualised using an Enhanced Chemi-Luminescence (ECL) system (Amersham Biosciences). Blots were stripped with 5% acetic acid for 5 minutes, and re-probed with mouse anti-GAPDH or actin (both Abcam) to confirm that equal amounts of protein had been

loaded. Blots were stripped up to three times for re-probing for other proteins of interest, or for optimisation of a new antibody.

2.9 cDNA constructs

CFP-actin and YFP-actin constructs were made by James Monypenny (Dunn et al., 2002). GFP-α-Actinin was a gift from Marina Guvakova (University of Pennsylvania). GFP-V14 Rho was a gift from Ferran Valderrama (UCL). All other constructs were produced as follows.

2.9.1 Candidate gene cloning

2.9.1.1 Strategy

The coding sequence (CDS) of each candidate gene was obtained from Genbank (http://www.ncbi.nlm.nih.gov/Genbank). Primers were designed to anneal to the ends of the CDS and to incorporate restriction sites to be used for directional cloning into a pEGFP vector (Clontech). C-terminally tagged GFP fusions of BK and 4.1B, and N-terminally tagged GFP fusions of CASK, 4.1B and RIL were generated.

2.9.1.2 Generation of template cDNA

RNA was extracted from the cells of interest as described for the microarray (Section 2.5.7.1). 10 μg of total RNA and 5 μM oligo(dT) primer (Affymetrix) were denatured by incubation at 65°C for 10 minutes, and then allowed to anneal by cooling on ice for 5 minutes. The RNA was then reverse-transcribed into cDNA using Superscript II Reverse Transcriptase (Invitrogen) and incubated at 42°C for 60 minutes. The cDNA was then treated with 2 units of RNase H (Ambion) for 20 minutes at 37°C. The cDNA was then cleaned-up using a Qiaquick PCR Purification Spin Column (Qiagen) according to the manufacturer's instructions, and eluted in 10 mM TrisCl, pH 8.5. The amount and purity of the cDNA were determined using readings taken at 260 nm and 280 nm on a Nanodrop spectrophotometer.

2.9.1.3 Primers

The sequences of the primers used for cloning into pEGFP-based vectors are shown below. The base pairs complementary to the mRNA sequence are shown in blue, and the enzyme recognition sequences are shown in red.

Primers used for C-terminal tagging with GFP

BK forward including *Xho I* site,
5'-AATTAACTCGAGGCAGCAATGGCGTACATCCAGTTGGAACCACTGAACG-3';
BK reverse including *Kpn I* site,
5'-ATTAAGGTACCATGCAGCATCAGGTCACCTCCAGCGAGGCCGGAGACAC-3'.

Primers used for N-terminal tagging with GFP

CASK forward including *Xho I* site,
5-AATTAACTCGAGGGAGCAATGGCCGACGACGACGTGCTGTTCGAGGAT-3';
CASK reverse including *Kpn I* site,
5'-TTAATTGGTACCTGCTGCCTAATAGACCCAGGAGACCGGGACCCACTG-3';
4.1B forward including *Hind III* site,
5'-TAAATCAAGCTTGCAGCAATGACAACCGAATCAGGATCAGACTCAGAA-3';
4.1B reverse including *Kpn I* site,
5'-TTAATTGGTACCTGCTGCTCAATCCTCTCCATCTTCTGGTGTGATTTC-3';
RIL forward including *Hind III* site,
5'-ATATATAAGCTTGCAGCAATGACCCACGCGGTGACCCTGCGCGGCCCT-3';
RIL forward including *Kpn I* site,
5'-GTGACTGGTACCTGCTGCTCAGACAAGTTCCACCTTGGCATTGGGATA-3'.

2.9.1.4 Polymerase Chain Reaction

For amplification of cDNA, KOD Hot Start DNA Polymerase (Novagen) was used. The components of each PCR reaction were as follows; 1 × PCR Buffer for KOD Hot Start DNA Polymerase, 1 mM $MgSO_4$, 0.2 mM dNTPs, 100 ng cDNA, 0.6 µM primer pair and 10% DMSO. Thermal cycling was performed on a PTC-200 Peltier thermocycler (MJ Research) with the following conditions; 2 minutes denaturation at 94°C, followed by 30 cycles of 15 seconds at 94°C, 30 seconds at 65°C, 2 minutes at 72°C, and a final extension step at 72°C for 5 minutes. The PCR product was electrophoresed on an agarose gel alongside a 1 kb DNA ladder (Invitrogen) and the correctly-sized band excised and purified using a Qiaquick PCR Purification Spin Column as before.

2.9.1.5 Restriction digestion and ligation of cDNA

The cDNA was double-digested overnight at 37°C, with the restriction enzymes whose recognition sequences had been included in the primers. The reaction was terminated by incubation at 75°C for 10 minutes and cleaned-up using a Qiaquick PCR Purification Spin Column. The digested CASK, RIL and 4.1B cDNAs were ligated into a similarly-cut pEGFP-C3 vector (Clontech) and BK cDNA was ligated into similarly-cut pEGFP-N1 vector (Clontech) overnight at 16°C using T4 DNA Ligase (NEB), with a vector:insert molar ratio of 1:4.

2.9.1.6 Transformation of E.coli and plasmid preparation

An aliquot of One Shot TOP10 Chemically Competent *E.coli* (Invitrogen) was transformed with the ligation reaction according to the manufacturer's instructions and plated onto LB Agar plates supplemented with 50 μg/ml Kanamycin A. Plates were incubated overnight at 37°C. The following morning, colonies were picked and inoculated into 3 ml LB supplemented with 50 μg/ml Kanamycin A, and incubated with shaking at 37°C for at least six hours. The samples were then centrifuged at 5000 rpm for 10 minutes and plasmid DNA was prepared from the resulting bacterial pellets using Qiagen's Miniprep system according to the manufacturer's instructions. Miniprepping was carried out by staff at the Cancer Research UK London Research Institute Equipment Park. Once the sequence of the clone had been verified, large plasmid stocks were purified using the QIAfilter Plasmid Maxi Kit (Qiagen) according to the manufacturer's instructions.

2.9.1.7 DNA sequencing

The sequences of the inserts were checked by performing PCR-cycle sequencing using the BigDye Terminator v3.1 system (ABI) and an ABI 377 automated sequencer (PE Applied Biosystems). Sequencing was carried out by staff at the Cancer Research UK London Research Institute Equipment Park. Typically the insert was sequenced with forward and reverse primers flanking the multiple cloning site of the GFP vector, as well as with internal, gene specific primers for reading sequences further away from the ends of the insert. Primers used for the sequencing of cDNA fragments inserted into the multiple cloning site of pEGFP-C3 were designed by James Monypenny (2003) and are; forward, 5'-CGGCATGGACGAGC-3', and reverse, 5'-GCTGCCATGGCGCCC-3'. Sequences of the gene specific primers used to read sequences far from the ends of the insert are, GFP-BK (downstream of base 771), 5'-GGAACGCTACACCTTCGAGA-3'; GFP-CASK (downstream of base 1125), 5'-CCAGGATCAGCATCTTCACA-3'; GFP-CASK (downstream of base 1500), 5'-TGAGCCAATGGGAATCACT-3'; GFP-4.1B (downstream of base 800), 5'-AACCACACGAAAGAACTGGA-3'; GFP-4.1B (downstream of base 724), 5'-GGGGACTATGATCCT-3'; GFP-4.1B (downstream of base 1327), 5'-ATGTCTCGCAGCTTGGATGGA-3'.

2.9.2 Subcloning

2.9.2.1 Generation of untagged cDNAs

Where GFP-fusion proteins were not required, the GFP gene was excised from the construct by digestion with *AgeI* and *BsrGI*. The digested vector was then treated with Mung Bean Exonuclease (NEB) for 30 minutes at 30°C to remove 5' overhangs. The product was cleaned-up and ligated overnight at 16°C, transformed into *E.coli* and mini-prepped as usual. Plasmids were analysed by agarose gel electrophoresis to check that the GFP had been removed, and then resequenced to ensure that the cDNA of interest was still present and in the correct reading frame.

2.9.2.2 Generation of C-terminally GFP-tagged 4.1B

A C-terminally tagged 4.1B construct was generated by amplifying 4.1B from the original N-terminally tagged construct (see Section 2.9.1) by PCR using the following primers, 4.1B forward including *Hind III* site,
5'-AAATCAAGCTTGGAGGAATGACAACCGAATCAGGATCAGACTCAGAA-3';
4.1B reverse including *Kpn I* site,
5'-TTAATTGGTACCAATCCTCGTCCATCCTCTCCATCTTCTGGTGTGATTTC-3'.

The DNA was then digested with *HindIII* and *KpnI* and ligated into a similarly-cut pAcGFP-N1 vector.

2.10Sequence analysis

2.10.1DNA sequence analysis

In the primer design process, DNA sequences were manipulated using the Sequence Manipulation Site (http://www.bioinformatics.org/sms/). Sequence trace data obtained by BigDye Terminator PCR-cycle sequencing were viewed with FinchTV (Geospiza, Inc). DNA sequences were aligned using the Clustal W program (Chenna et al., 2003). DNA sequences were translated into protein sequences using the Sequence Manipulation Site.

2.10.2Protein sequence analysis

Protein sequences of the *epb41l3* splice variants were superficially investigated by searching for protein domains with Pfam (Bateman et al., 2004). Predicted domains with E-values < 0.05 were accepted.

2.11RNA interference

RNA interference was achieved using siRNA oligos (Ambion) introduced to the cells by lipofection or microinjection, depending on the requirements of the particular experiment. Annealed purified siRNA oligos were from Ambion, Inc. and were resuspended according to their protocol. The siRNA sense sequences are as follows,
Rat *actn1*, GGAUGGUCUUGGUUUCUGU;
Rat *cask*, GCGGGAAGCCAGUAUCUGU;
Rat *epb41l3*, GCAUGCAGUGCAAAGUGAC;
Human *epb41l3* oligo 1, GCAUCACUAAACCGAUAAU;
Human *epb41l3* oligo 2, GCUCGAAUAUCAGCAAUUA;
Human *epb41l3* oligo 3, GCGAUUACAUUAGUGAGUU.
The control was Silencer Negative Control #1 siRNA (sequence not made available by Ambion, Inc.). Oligos were labelled with Cy3 using Silencer siRNA Labelling Kit (Ambion, Inc.) for studies where it was important to be certain of exactly which cells contained siRNA. The efficacy of the siRNA at reducing expression of the target protein after 48 hours was first checked by immunoblotting, and sometimes

immunocytochemistry. A negative control siRNA was performed in all RNAi-based experiments. Cells were used in experiments 48 hours after transfection.

For RNAi rescue experiments, Hela cells were transfected with rat 4.1B cDNA 48 hours after siRNA transfection, and the effects of the rescue assessed after 24 hours (72 hours after the first transfection). The persistence of RNAi after 72 hours was also checked by immunoblotting. RNAi rescue was achieved because the three Human *epb41l3* siRNA oligos targeted sequences that were not present in the cDNA for rat *epb41l3*.

2.12 Transfection of nucleic acids into cultured cells

2.12.1 Microinjection

Cells seeded on 22 mm × 22 mm, No. 1.5, coverslips were microinjected with an Eppendorf transjector and micromanipulator system mounted on a Zeiss Axiovert 35 equipped with a tungsten lamp, a long-working distance condenser and a ×20, NA 0.3, long-working distance objective. Microinjection needles were made from glass capillaries (Fisher Scientific) using an electronic microfilament puller (P-97, Sutter Instrument Co.). Solutions were prepared for microinjection by centrifugation at 14,000 rpm for at least 20 minutes in order to sediment particles that would have blocked the microinjection needle. siRNA oligos were microinjected into the cytoplasm at a concentration of 2 μM, and cDNA constructs were microinjected into the nucleus at a concentration of 0.05 μg/μl.

2.12.2 Lipofection

Cells were seeded at 40-50% confluence the day before transfection. Chemical transfection was performed using Effectene (Qiagen) according to the manufacturer's instructions. For cDNA transfection, 0.3 μg/ml DNA was used and for siRNA transfection, 100 nM siRNA oligos were used. Briefly, for one well of a 12-well plate, ~30,000 cells were seeded and the following day the cells were washed and 800 μl fresh medium added. The nucleic acids were then added to 72 μl of EC buffer (provided with the Effectene reagent), and 2.5 μl Enhancer solution was added. The tube was then vortexed briefly and incubated at room temperature for 2 minutes. 6 μl of Effectene was then added and the tube was vortexed for 10 seconds and incubated at room temperature for 5 minutes. 400 μl medium was then added and mixed by pipetting up and down gently, and the mixture then added to the cells. The day after transfection the cells were

washed and fresh medium was added to the cells. Expression of cDNA was usually seen after 12 hours, and a reduction in protein expression caused by siRNA transfection usually occurred after 48 hours.

Chapter Three

Characterisation of the rat sarcoma model of metastasis

Metastasis, the dissemination of cancerous cells from a primary tumour leading to the growth of a secondary tumour at a distant site, is the most life-threatening aspect of cancer. Therefore, the mechanisms of metastasis are of great interest. My approach to studying the mechanisms of metastasis was based upon a rat sarcoma model composed of four cell lines with different abilities to metastasise.

I began by characterising the cell populations comprising the model. I focussed on cell motility and the cytoskeleton, since previous work (Pokorna et al., 1994) had indicated that other cell populations from this model have distinct F-actin cytoskeletal arrangements. Moreover, remodelling of the actin cytoskeleton is essential for cell migration and invasion, which play a role in metastasis. I also looked at the arrangement of F-actin by imaging fixed cells stained with rhodamine phalloidin.

Finally, I investigated cell motility, in terms of the chemotactic responses of the cells to a gradient of PDGF and IGF. These growth factors were chosen since a related sarcoma cell population was already known to chemotax towards PDGF/IGF, furthermore, chemotaxis towards blood vessels and organs may be important in the metastatic cascade.

3.1 The actin cytoskeleton

3.1.1 Differences in F-actin organisation between the sarcoma cell populations

Differences in F-actin organisation between the non-metastatic and metastatic cells were observed after staining fixed cells with rhodamine phalloidin, and imaging cells on a confocal microscope. The images in Figure 3.1 show the differences in F-actin organisation. The non-metastatic K2 cells tend to have F-actin arranged into thick actin stress fibres, which sometimes traverse the cell. These cells are generally well spread. In the metastatic T15, A297 and A311 cells, on the other hand, stress fibres occur less frequently. They are usually polarised and their actin is concentrated into ruffles at the leading edge, and elsewhere arranged into a network of disordered filaments.

The non-metastatic K2 cells have significantly more stress fibres than their metastatic counterparts T15, A297 and A311. Quantification of randomly-acquired fields revealed that 68% of K2 cells contain at least one stress fibre, in comparison to only 7-10% of the metastatic cells (Figure 3.2).

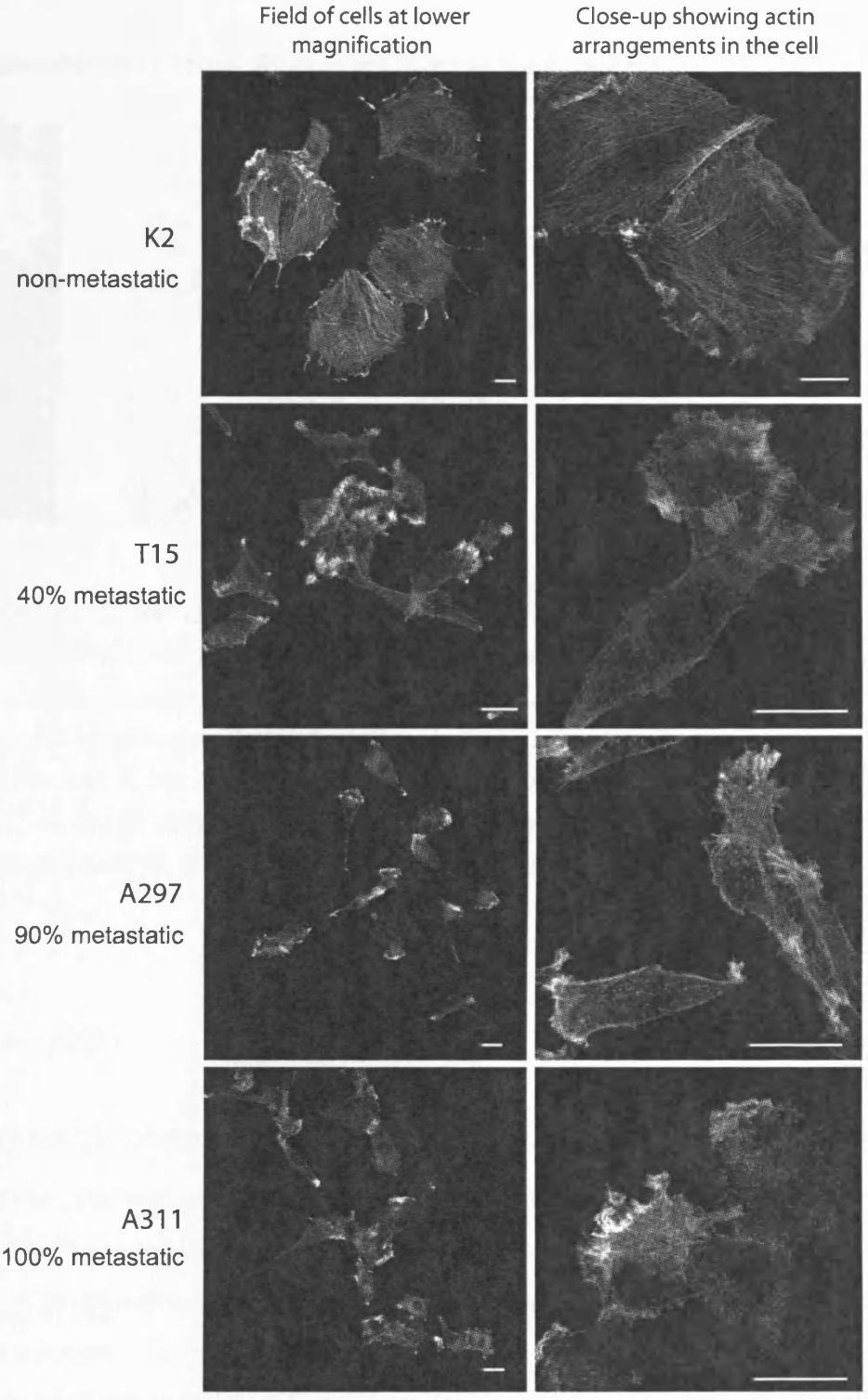
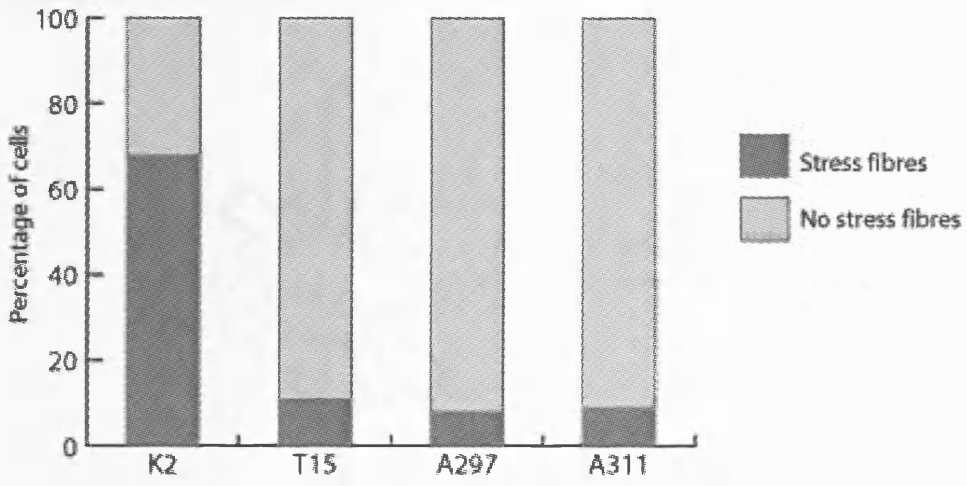


Figure 3.1: F-actin organisation in the four cell populations.
These confocal images of rhodamine phalloidin-stained cells show the distinct arrangements of F-actin in the non-metastatic and metastatic cells. F-actin in the K2 (non-metastatic) cells is generally arranged into stress fibres, which sometimes traverse the cell. The K2 cells tend to be larger than the other three populations, and well spread. F-actin in the metastatic cell populations (T15, A297 and A311) is typically arranged into ruffles at the leading edge of the cell, and elsewhere, into a disordered network of thin filaments. These cells spread less than the K2 cells and are usually polarised. Images in the right panel were taken at high magnification for optimal observation of F-actin structures. Images in the left panel were taken at a lower magnification to demonstrate that the phenotypes are representative across a larger field of cells. Scale bar = 20 μm.

Figure 3.2: Distribution of stress fibres in the four cell populations.



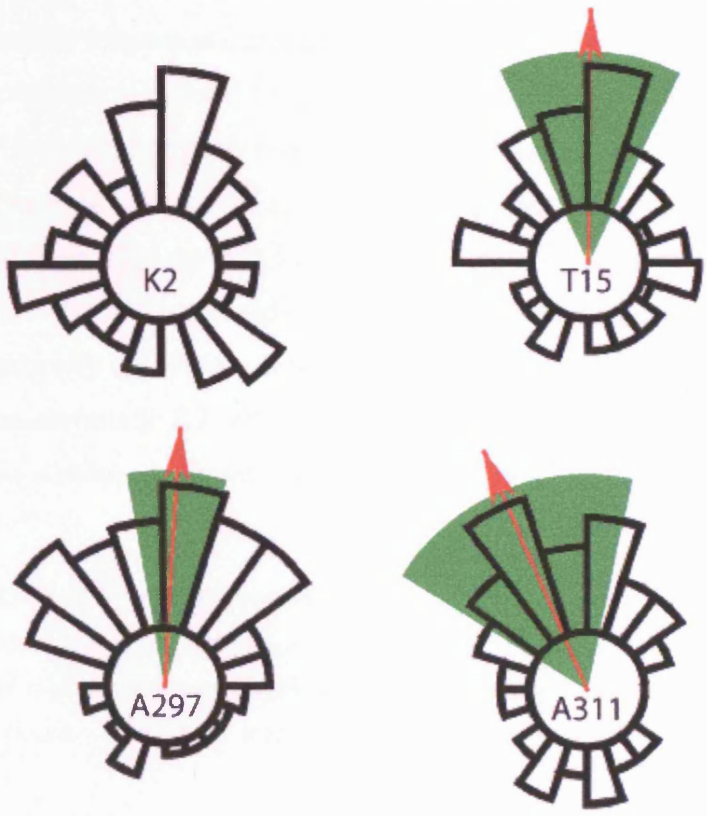
This graph shows the percentage of actin stress fibre-containing cells in each population of the metastasis model. These data were obtained from randomly selected fields of rhodamine phallodin-stained cells acquired on a confocal microscope. Images were subjected to a blind analysis, in which a cell was classified as having stress fibres if it contained at least one actin fibre that could be seen at a magnification of ×25. The numbers of cells scored for K2, T15, A297 and A311 are 50, 55, 40 and 55 respectively. The difference in stress fibre distribution between the K2 and the T15, A297 and A311 is significant (χ^2 $P < 0.005$).

3.2 Cell motility

3.2.1 Chemotactic responses of the cells to PDGF/IGF

Responses of the cells to a gradient of PDGF/IGF were tested in the Dunn chemotaxis chamber and are illustrated in Figure 3.3. The circular histograms present the distributions of the directions of the cell tracks and illustrate the differences in the chemotactic responses. The K2 cells are not chemotactic; in contrast, the T15, A297 and A311 cells have strong and significantly greater chemotactic responses. These data represent at least five independent experiments for each cell population. Only cells that had migrated further than 70 μm were included in the evaluation of chemotaxis.

Figure 3.3: Chemotactic responses of the sarcoma cell populations to PDGF/IGF.



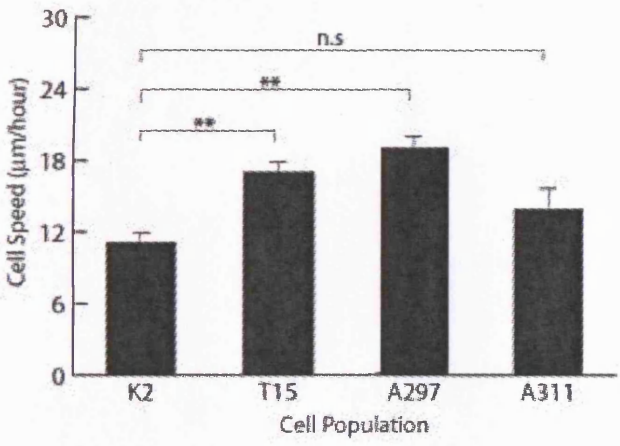
The circular histograms represent the chemotactic responses of the cell populations to a concentration gradient of 60 ng/ml PDGF and 80 ng/ml IGF in the Dunn chemotaxis chamber over 16 hours. Here, the well containing the growth factors is at the top of the histogram, and the growth factors diffuse towards the well containing the starvation serum, at the bottom of the histogram. The red arrow indicates the mean direction of cell migration, and the green wedge indicates its 95% confidence interval. The non-metastatic K2 cells are not chemotactic to the gradient of PDGF/IGF, but the metastatic T15, A297 and A311 cells have a strong chemotactic response. The chemotactic responses of the K2 and T15/A297/A311 cells are significantly different (ANOVA $P < 0.05$). These data represent at least five independent experiments, and the total numbers of cells analysed for the K2, T15, A297 and A311 cell populations are 61, 81, 86 and 32 respectively.

3.2.2 Differences in the speed of migration between the cell populations

The cell populations migrate at different speeds in the Dunn chemotaxis chamber (Figure 3.4). The speeds were calculated from the same data used for the evaluation of chemotaxis, without applying any restriction on the distance moved, meaning that all the cells recorded were included in the analysis. The non-metastatic K2 cells move at a mean speed of 11 μm/hour, and the T15, A297 and A311 cells move at 17, 19 and 14 μm/hour respectively. The increased speed in the T15 and A297 cells compared to the K2 cells is statistically significant. However, there is no significant difference in speed between the non-metastatic K2 cells and the most metastatic, A311 cells, in fact, the A311 cells show a reduced migration compared to the T15 and A297 cells.

Figure 3.4: Speeds of in vitro migration of the sarcoma cell populations.

*Cells were exposed to a concentration gradient of PDGF/IGF in the Dunn chemotaxis chamber. A low light level digital microscope was used to acquire one phase contrast image every 5 minutes for 16 hours. Cells were tracked and the speeds evaluated using a Mathematica® Notebook developed in the laboratory. The mean speed of migration of the non-metastatic K2 cells is 11 μm/hour, whereas the metastatic T15 and A297 cells migrate significantly faster, at 17 and 19 μm/hour. The difference in speed between K2 and A311 is not significant. The total numbers of cells analysed for the K2, T15, A297 and A311 cell populations are 171, 116, 177 and 74 respectively. Error bars represent the SEM. Statistical significance is indicated as follows; **P < 0.05; n.s = not significant.*



A more detailed analysis revealed that the cells within each population are heterogeneous and that subpopulations of fast-migrating cells exist. For example, in the A311 population, of which the average speed of migration is 14 μm/hour, there is great heterogeneity and some cells migrate at a speed of up to 40 μm/hour (see bottom histogram in Figure 3.5A). In comparison, the K2 cells, which have an average speed of 11 μm/hour, are less heterogeneous and contain fewer fast-migrating cells. When the fractions of cells migrating faster than 20 μm/hour are compared, it becomes clear that the metastatic cells have a significantly larger fast-migrating fraction of cells (Figure 3.5B). The fast-migrating subpopulation of the non-metastatic cells K2 comprises 8% of the total K2 cell population, whilst the fast-migrating fractions of the T15, A297 and A311 cells make up 32, 34 and 20% respectively of the total cell populations.

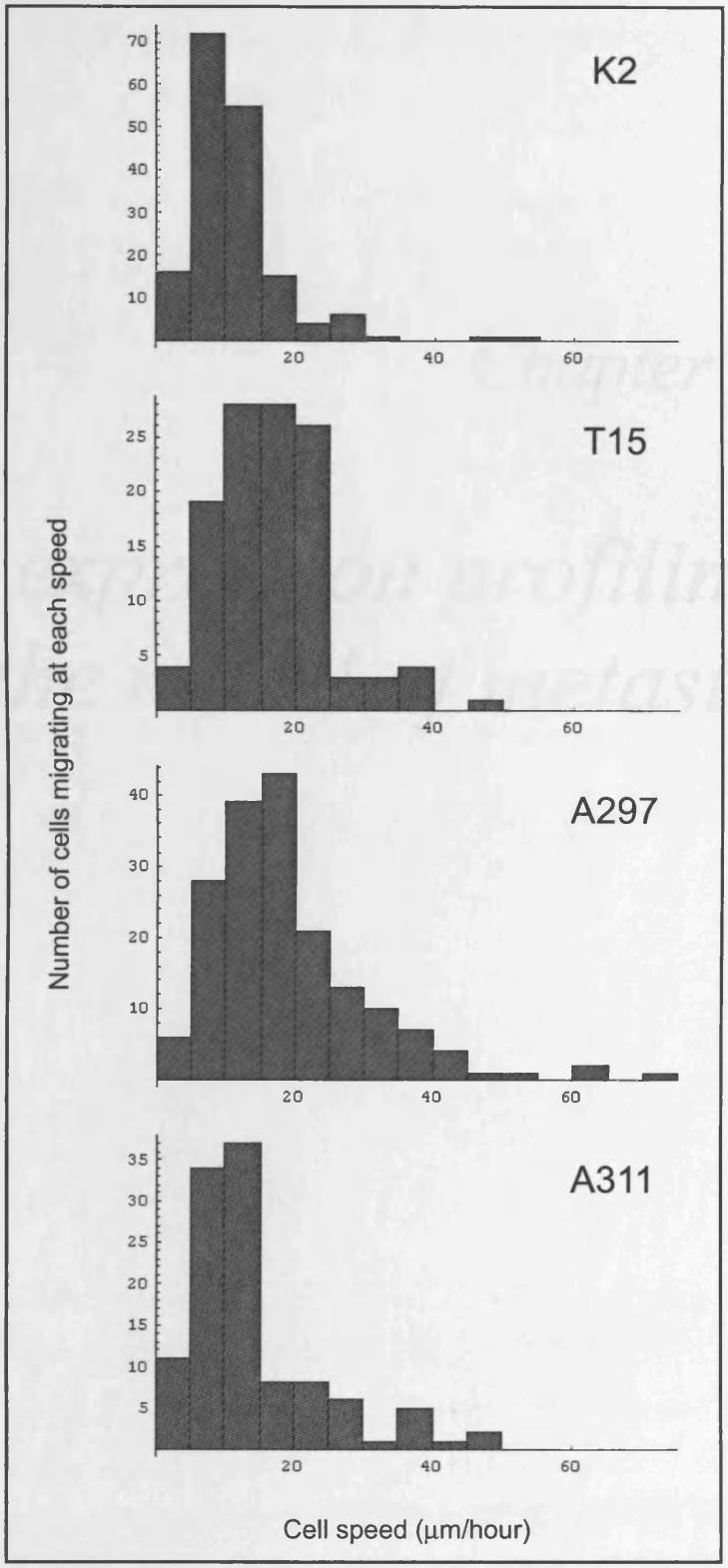
Figure 3.5: Distribution of cell speeds within the four populations.

A. The histograms illustrate the different distributions of speed between the four cell populations, and that within each, there are subpopulations of cells.

B. The proportion of fast-migrating cells (cells migrating faster than 20 μm/hour) is significantly increased in the metastatic compared to the non-metastatic populations (χ^2 $P < 0.01$).

Figure 3.5: Distribution of cell speeds within the four populations.

A



B

Cell population	Percentage of cells moving faster than 20 μm/hour
K2	8
T15	32
A297	34
A311	20

Chapter Four

Gene expression profiling of the model of metastasis

The previous chapter described differences between the cell populations in the metastasis model, which are summarised in the following table.

Table 3: Summary of differences between the sarcoma cell populations.

Cell type / Characteristic	K2	T15	A297	A311
Metastatic potential	0%	40%	90%	100%
F-actin cytoskeleton	Rich in stress fibres	Mainly arranged into a network of thin filaments; contains fewer and finer stress fibres		
Actin dynamics	Slow	Faster	Not tested	Not tested
Response to IGF/PDGF	Not chemotactic	Strongly chemotactic		
Speed of migration mm/hour	11	17	19	14

Some of these characteristics may correspond to the differences in metastatic potential. I compared the gene expression between the sarcoma cell populations with a view to identifying the genes responsible for these differences. Gene expression was measured using Affymetrix GeneChip® Rat 230A microarrays, which contain probes to measure the expression of most of the genes in the rat genome.

Two related microarray experiments were performed. The first experiment compared gene expression in K2 and A297 cells and their responses to PDGF/IGF. The second experiment included the cell populations T15 and A311, as well as tumours generated by implanting the cells subcutaneously into rats. Using the data obtained from both experiments, I looked for differences in gene expression between the non-metastatic and metastatic cell populations, filtering out genes whose expression did not correlate between cultured cells and primary tumours, and obtained a list of twenty-three candidate genes.

4.1 *Initial microarray experiment*

4.1.1 List of candidate genes

Affymetrix Rat 230A GeneChips® were used to compare the gene expression of the K2 and A297 cell populations, when subjected to (1) starvation in 0.5% serum for five hours, (2) starvation followed by 30 minutes treatment with PDGF/IGF (3) starvation followed by 3 hours treatment with PDGF/IGF. These responses will be referred to as control, early and late, respectively.

About two thirds of the genes on the Affymetrix GeneChip® had detectable expression levels in the cells. 294 genes (out of ~15,000 probed transcripts) were significantly differentially expressed between A297 and K2 cells, where a significant difference in gene expression was defined as an expression ratio > 1.6 ($P < 0.05$). The expression of these genes was different between populations, but unchanged by stimulation of the cells with PDGF/IGF. This is because the gene expression responses to the growth factors were relatively small, and were being masked by the relatively high number of differences between the cell populations.

I was interested in both the metastatic potential and the response to PDGF/IGF, since the cells are chemotactic to these growth factors. Therefore I applied different expression ratio (ER) and significance criteria to each dimension of the experiment. Changes in gene expression between the populations were defined by stringent parameters ($ER \geq 1.6$, $P < 0.05$) and changes in gene expression in the growth factor response were defined by less stringent parameters ($ER \geq 1.3$, $P < 0.1$), and a list of 298 candidate genes was produced (Appendix 3).

In this list, 95 genes had increased expression in metastasising cells and 157 genes had decreased expression in the metastatic cells. In the response of the metastatic cells to growth factors, 39 genes had differential expression. In the response of the non-metastatic cells to the growth factors, 25 were changed. A total of 16 genes exhibited differential expression both between the cell populations and in response to growth factors. These observations are expressed in Figure 4.1.

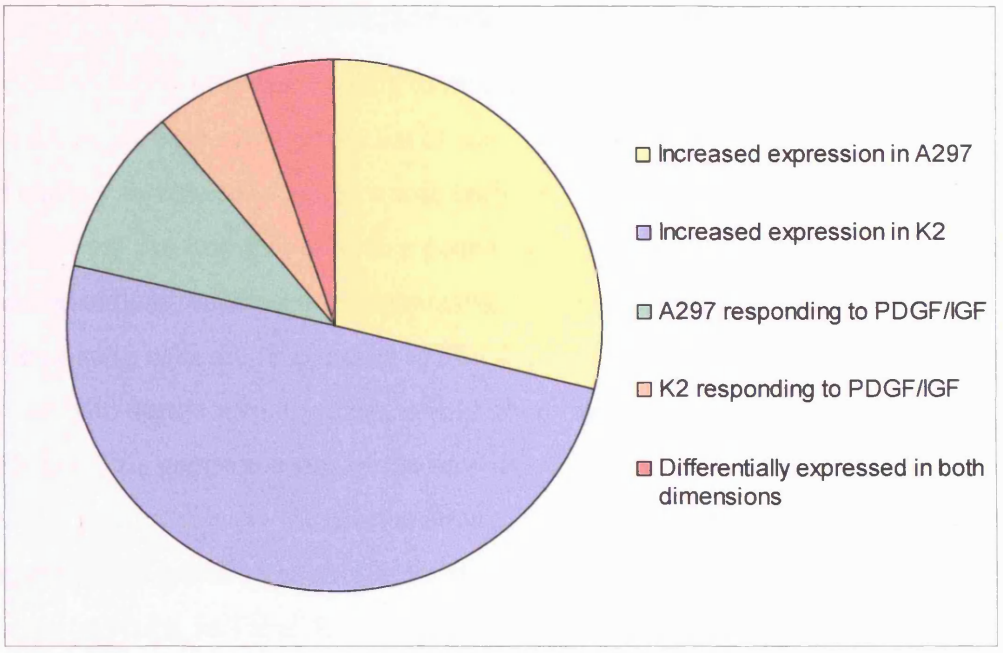


Figure 4.1: Summary of the gene expression patterns observed in the initial microarray experiment.

The candidate genes were mainly differentially expressed between the two cell populations, with almost half the genes being downregulated with metastasis. Some genes changed in response to growth factors, 16 of which were also differentially expressed between the two cell populations (represented by the red segment).

4.1.2 Candidate genes selected after manual inspection of the data

The list of 298 genes was too long to investigate fully, so I manually inspected the data in order to obtain a manageable list of candidate genes for further study. I was particularly interested in genes whose early or late responses were different between cell populations. An example of such a gene is microtubule-associated protein 6 (*mtap6*), which is not only increased in metastasising cells, but is also increased when metastasising cells are responding to PDGF/IGF. Such patterns may indicate genes that can mediate motile responses that would be of benefit to a metastasising cell. I also considered the genes in terms of the novelty of their connection with metastasis, and possible relationships to the mechanisms of metastasis, inferred from the literature. In this way, eight genes were chosen for further analysis and are presented, with their expression ratios, in Table 4.

Gene (Gene symbol)	Gene ontology: Biological process	Expression ratio AE/AC	AL/AE	KE/KC	KL/KE	AC/KC	AE/KE	AL/KL
Alpha actinin 1 (actn1)	Actin binding Cytoskeleton Calcium binding	1.1	0.9	1.1	1	1.5**	1.5	1.3
Bone Morphogenetic Protein (bmp2)	Skeletal development Epithelial to mesenchymal transition Transforming growth factor beta receptor	1.1	1.4	1.6	0.6	2.8*	1.9*	4.3**
Translocating chain-associating membrane protein (RGD:1359100)	None assigned	1	1.4	0.9	1.1	0.2**	0.2	0.2**
Matrix metalloproteinase 2 (mmp2)	Proteolysis Collagen catabolism Peptidoglycan metabolism	1	0.9	1	1	0*	0	0**
Matrix metalloproteinase 3 (mmp3)	Proteolysis Collagen catabolism Peptidoglycan metabolism	1.1	1	1.1	1.1	5*	5.1	4.4**
Microtubule-associated protein 6 (mtap6)	Microtubule-based process	1	1.4*	0.9	1.4	1.2	1.4	1.4*
Roundabout homolog 1 (robo1)	Chemotaxis Cell adhesion Homophilic cell adhesion	0.5	1.6	1	1	4	2.1	3.5**
Tropomyosin 1 alpha (tpm)	Muscle contraction Regulation of muscle contraction Muscle development	0.9	0.9	1.1	1.1	0.5*	0.4**	0.3

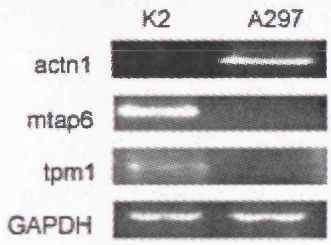
Table 4: Initial microarray experiment: Candidate genes selected after manual inspection of the data.

*Expression ratios represent expression compared between two samples. A=A297; K=K2; C=Control (5 hour starvation); E=Early response (starvation followed by 30 minutes IGF/PDGF); L= Late response (starvation followed by 30 minutes IGF/PDGF). The statistical significance of each expression ratio was calculated using the t-test and corrected with the Benjamini-Hochberg algorithm; * P=0.1, ** P=0.05, *** P=0.01. The principle three gene ontology terms specified for the biological process of the gene by Affymetrix are listed here.*

4.1.3 Validation of candidate genes by RT-PCR

Microarray data requires validation by at least one other method (Chuaqui et al., 2002). Therefore, some of the candidate genes were tested by RT-PCR to confirm the expression indicated by the microarray analysis. Figure 4.2 shows the RT-PCR products of *actn1*, *mtap6* and *tpm1*. The expression patterns of *actn1* and *tpm1* reflect those detected by the microarray. However, *mtap6* appears to be up-regulated in K2 cells, although the microarray data suggetsed that it is down-regulated in these cells. The microarray data were not always consistent with the RT-PCR data and microarray data were therefore only trusted when supported by RT-PCR.

Figure 4.2: Initial microarray experiment: Validation of gene expression by RT-PCR



RT-PCR was performed to check the expression of selected candidate genes. Total RNA was extracted from each cell population and reverse-transcribed into cDNA, and regions of each gene of interest were amplified by PCR. PCR products were resolved on a 1% TBE agarose gel supplemented with ethidium bromide, and visualized by UV transillumination. Tpm1 and actn1 are upregulated in A297 and K2 respectively, which reflects the gene expression pattern detected by the microarray. However, RT-PCR showed that mtap6 is upregulated in K2 cells, although the microarray data indicated that it is downregulated in these cells. RT-PCR for GAPDH was performed as a control.

4.2 Extended microarray experiment

4.2.1 Genes differentially expressed between metastatic/non-metastatic sarcoma cells and tumours

This experiment included all four cell populations as well as the primary tumours which arose after subcutaneous injection of these cells into the inbred rats. The extended dataset was processed in the same way as the initial experiment. Metastatic cells were compared to non-metastatic cells with six comparisons being made; T15 cells vs. K2 cells, A297 cells vs. K2 cells, A311 cells vs. K2 cells, T15 tumours vs. K2 tumours, A297 tumours vs. K2 tumours, and A311 tumours vs. K2 tumours. The criteria applied to generate a list of genes were firstly, that each of the six expression ratios should be greater than 2, and secondly, that five out of the six expression ratios should be significant ($P < 0.05$). The resulting list of eighty genes is shown in Table 5.

Table 5: Extended microarray experiment: The eighty most differentially expressed genes between non-metastatic and metastatic cells and tumours.

*In this table, the Affymetrix Identifiers, expression ratios and symbols for each differentially expressed transcript are displayed. The Affymetrix Identifier is the unique identification for each transcript that is probed by the array. Not all transcripts are yet characterised, which is why only forty-four of the Affymetrix Identifiers in this table have accompanying gene symbols. Where the gene's identity has been inferred from sequence similarity to another gene, the gene symbol is enclosed in parentheses. The middle columns describe the expression ratio of two experimental groups, where t = T15 cells, a = A297 cells, k = K2 cells, m = A311 cells, tt = T15 tumour, at = A297 tumour, kt = K2 tumour, mt = A311 tumour. High expression ratios (ER > 1) indicate that the gene is over-expressed in metastatic cells. Low expression ratios (ER < 1) indicate that the gene is under-expressed in metastatic cells. Statistical significance is indicated as follows; *P < 0.1, **P < 0.05, ***P < 0.01.*

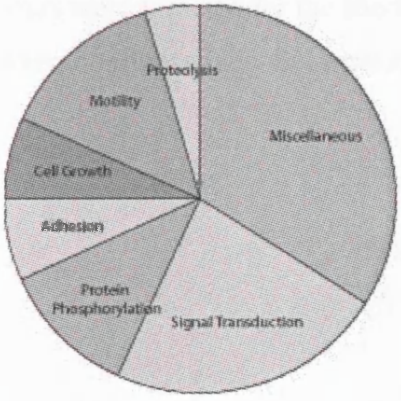
Affymetrix Identifier	t/k	a/k	m/k	tt/kt	at/kt	mt/kt	Gene symbol
1369113_at	0.***	0.***	0.***	0.**	0.**	0.**	Cktsf1b1
1372820_at	0.***	0.***	0.***	0.1*	0.**	0.**	
1370696_at	0.***	0.***	0.***	0.**	0.*	0.**	LOC207121
1372539_at	0.***	0.***	0.***	0.**	0.1*	0.**	
1377029_at	0.***	0.***	0.***	0.***	0.**	0.**	
1370503_s_at	0.***	0.***	0.**	0.1**	0.1***	0.***	Epb4.1l3
1368955_at	0.**	0.***	0.***	0.1**	0.1***	0.1**	Cask
1372809_at	0.1***	0.***	0.***	0.***	0.***	0.***	(AF152002)
1376071_at	0.1***	0.***	0.1***	0.3**	0.2**	0.2**	
1374318_at	0.2***	0.***	0.***	0.5**	0.2***	0.2**	
1368106_at	0.1**	0.***	0.1***	0.2***	0.2***	0.2**	Plk2
1369972_at	0.2**	0.***	0.1**	0.3	0.1**	0.1**	Serpinb5
1382117_at	0.3**	0.***	0.***	0.4	0.1**	0.***	(FLJ21148)
1369126_at	0.3*	0.1**	0.1**	0.**	0.**	0.**	Ptgfr
1376708_at	0.2***	0.1**	0.1**	0.3**	0.4**	0.4*	
1375986_at	0.2***	0.1***	0.1***	0.1***	0.1**	0.1***	
1371475_at	0.2***	0.1***	0.1***	0.4***	0.2***	0.1**	Rnase4
1387653_at	0.3***	0.1***	0.1**	0.4**	0.3**	0.2***	Tsnax
1374953_at	0.2***	0.2***	0.1***	0.2**	0.2**	0.1***	IMAGE:7110106
1368533_at	0.4**	0.***	0.1***	0.3**	0.2***	0.1***	Heph
1398823_at	0.3**	0.1**	0.1***	0.4***	0.4***	0.3**	Tsnax
1388821_at	0.3**	0.1***	0.1***	0.6**	0.1***	0.1***	(AW319517)
1368943_at	0.3**	0.2***	0.1***	0.5**	0.2***	0.2**	Rnase4
1377181_at	0.2**	0.2*	0.3**	0.3**	0.2***	0.3**	
1368223_at	0.2**	0.3**	0.2**	0.3**	0.1***	0.1**	Adamts1
1369652_at	0.2***	0.2***	0.3***	0.5*	0.4**	0.4**	Thy1
1390638_at	0.5**	0.***	0.1**	0.3**	0.1**	0.1*	
1372310_at	0.2***	0.2**	0.2***	0.5**	0.6**	0.4**	(trp)
1376047_at	0.4***	0.1***	0.1***	0.4*	0.1**	0.1**	
1389712_at	0.5**	0.1***	0.2***	0.1**	0.1***	0.1***	
1368021_at	0.4**	0.2***	0.2**	0.3**	0.3**	0.4**	Adh1
1377506_at	0.3**	0.2**	0.2**	0.1**	0.1**	0.1**	(GDF-1)
1391428_at	0.1***	0.3**	0.4**	0.1**	0.1**	0.1**	
1374652_at	0.4**	0.1***	0.3***	0.2**	0.2**	0.4	
1375729_at	0.5**	0.1**	0.2**	0.4*	0.2**	0.2**	
1372655_at	0.3***	0.2***	0.3***	0.4**	0.4**	0.4**	
1374728_at	0.3**	0.2**	0.3**	0.3**	0.3**	0.4*	
1376786_a_at	0.3***	0.2***	0.3***	0.3***	0.3***	0.3**	
1398363_at	0.4**	0.2**	0.2***	0.8	0.3**	0.3**	
1390468_at	0.4***	0.2***	0.4***	0.5**	0.5**	0.5**	
1373198_at	0.5**	0.3**	0.3**	0.4***	0.3**	0.2**	(2810451A06)
1374817_at	0.4**	0.4*	0.3**	0.1***	0.1***	0.1***	
1370642_s_at	0.4**	0.4**	0.3***	0.7	0.4**	0.5***	Pdgfrb
1374743_at	0.5***	0.2***	0.5***	0.4**	0.4**	0.5*	
1374448_at	0.4***	0.4**	0.4**	0.4**	0.5*	0.5**	(invasion & met)
1374685_at	0.5**	0.4**	0.3***	0.3**	0.1**	0.2**	
1372870_at	0.5**	0.4***	0.4***	0.5**	0.4**	0.3***	

Affymetrix Identifier	t/k	a/k	m/k	tt/kt	at/kt	mt/kt	Gene symbol
1381012_at	0.4**	0.4**	0.4**	0.5**	0.5**	0.4**	
1388628_at	0.4***	0.4***	0.5***	0.5**	0.4**	0.3***	MGC94283
1367967_at	0.4**	0.5**	0.5**	0.4**	0.3***	0.3**	Lepre1
1372469_at	0.6*	0.4***	0.4**	0.4***	0.3***	0.4***	
1386969_at	2.**	2.2**	1.3	4.4**	3.4**	3.5**	Nrn1
1374066_at	2.1**	2.3**	2.3**	2.1**	1.9	2.2**	
1372650_at	2.2**	2.3**	2.2**	1.8**	2.5**	2.**	
1372069_at	2.1**	2.6**	2.1**	3.**	3.2***	3.7***	(bA130C19.2)
1370398_at	2.2**	2.4**	2.2**	2.2**	2.2**	1.9*	Wmp1
1372935_at	1.7*	3.**	2.1**	2.5**	3.4***	3.4**	
1387750_at	1.9**	2.6**	2.4**	2.6**	2.7**	2.4**	Twist
1375855_at	3.1***	2.7***	2.2**	5.6***	2.9**	2.1	
1375767_at	2.3**	3.4***	3.1***	1.6*	2.4**		
1390474_at	2.9***	2.9***	3.2***	3.6**	2.8***	3.5***	(9330161F08)
1373494_at	2.5**	4.***	2.6**	2.5**	2.2**	3.**	(bcr)
1370942_at	2.7***	3.9**	2.7***	2.4**	2.6**	2.9*	(Rasa3)
1388718_at	2.8**	4.9**	2.4*	6.5**	7.5**	7.4**	Tmod1
1398305_at	2.7**	3.1**	4.5**	2.3**	3.4**	2.9**	Bk
1369249_at	4.1**	4.7**	3.6**	3.7**	3.9*	3.7**	Ank
1387153_at	4.1***	4.5***	4.9***	3.1**	3.5**	3.9**	Ril
1373617_at	4.3***	5.2***	4.2***	3.1**	2.7**	2.6*	
1370418_s_at	3.6***	4.2***	5.9***	3.5***	4.7***	4.6***	Bk
1376304_at	3.8**	5.4**	4.7**	2.5**	3.1**	2.2**	
1373642_at	5.7***	3.7***	6.1***	3.6**	4.3***	4.1**	
1370638_at	6.1***	4.3**	6.***	11.8**	11.6**	16.3**	Ank3
1377065_at	5.8**	6.4***	5.3**	10.1**	12.7**	12.7**	
1371363_at	5.2***	9.3***	8.7***	1.	2.3**	2.6**	
1368342_at	5.2**	13.6***	10.2***	3.5***	4.1**	5.7*	Ampd3
1368713_at	8.2**	26.1**	3.9	17.**	18.8**	42.***	Mmp10
1371679_at	18.7***	15.2***	17.4***	9.3***	8.2**	8.3***	
1371132_a_at	18.1***	17.3***	21.3***	13.6**	21.7**	24.1**	Ank3
1370987_at	59.1***	42.6***	49.6***	2.2**	2.6**	2.3**	Spn
1368685_at	54.***	76.6***	64.9***	12.8***	14.6***	15.3***	Cspg4

4.2.2 Functions of the differentially expressed genes

Forty-four of the top eighty genes had been assigned a gene ontology (GO) term describing their biological process. The genes were grouped into categories according to their GO term. The main categories of these genes were signal transduction, motility, protein phosphorylation, cell growth, adhesion and proteolysis. Some GO terms did not readily fit into a category, or form a new category, and were grouped together as "miscellaneous". These data are displayed to give an impression of the functions of some of the most differentially expressed genes (Figure 4.3).

Figure 4.3: Extended microarray experiment: Functions of the eighty most differentially expressed genes.



Forty-four of the eighty most differentially expressed genes had been assigned GO terms describing their biological process. These were grouped into seven categories, which are displayed to give an impression of the functions of the most differentially expressed genes. Terms featuring in the list of the top eighty genes are adhesion, cell growth, cell motility, protein phosphorylation, proteolysis and signal transduction. Some terms (such as organ morphogenesis or mRNA cleavage) did not readily fit into a category or form a category of their own, and were grouped together as "miscellaneous".

4.2.3 Shortlist of candidate genes

Genes differentially expressed between non-metastatic and metastatic cell populations were shortlisted if they had an expression ratio > 2.5 ($P < 0.05$). In order to ensure that the gene expression differences observed were not solely due to the effects of cell culture, we compared gene expression in the cultured cells with expression in the corresponding primary tumours. We filtered out genes whose *in vitro* and *in vivo* patterns of expression were significantly different. This generated a list of twenty-three genes (Table 6). Of these genes, fifteen are down-regulated in metastasising cells and eight are up-regulated in metastasising cells. Graphs of gene expression were produced for each candidate gene and are shown in Appendix 4. Figure 4.4 shows the gene expression graphs for *bk, cask, epb41l3* and *ril*, which are analysed in more detail in the next chapter, since they were thought to be suitable candidates for functional assays. The graphs illustrate that only genes whose expression correlated between cells and tumours were selected for the shortlist, and that there are differences in gene expression between non-metastatic and metastatic cells.

Table 6: Extended microarray experiment: Genes with greatest differential expression between non-metastatic and metastatic sarcoma cells and tumours.

Gene downregulated in metastatic cells		
Gene (gene symbol)	Gene ontology (biological process)	Expression ratio
Gremlin 1 homolog (Grem1)	Cell-cell signalling; embryonic patterning; organ morphogenesis	0.006
RAR-related orphan receptor alpha (Rora_predicted)	Transcriptional regulation; DNA-dependent signal transduction	0.027
Erythrocyte protein band 4.1-like 3 (Epb41l3)	Cell proliferation; cortical actin cytoskeleton; actin organisation	0.028
Ca/CaM-dependent serine protein kinase (cask)	Protein complex assembly; protein phosphorylation; transport	0.031
Hypothetical gene AF152002 (LOC290595)		0.063
EST; Affymetrix Identififer 1376071_at		0.069
Polo-like Kinase 2 (Plk2)	Protein phosphorylation; cell cycle	0.080
EST; Affymetrix Identifier 1375986_at		0.14
RNase A family 4 (Rnase4)	mRNA cleavage	0.15
Similar to CG12279-PA (LOC500420)		0.17
Translin-associated factor X (Tsnax)		0.18
A disintegrin-like and metalloproteinase (Adamts1)	Proteolysis; integrin signalling; cell proliferation	0.22
Growth differentiation factor 1(Gdf1_predicted)		0.25
EST; Affymetrix Identifier 1391428_at		0.25
Neogenin (Neo1)	Protein phosphorylation; cell motility; cell adhesion	0.29

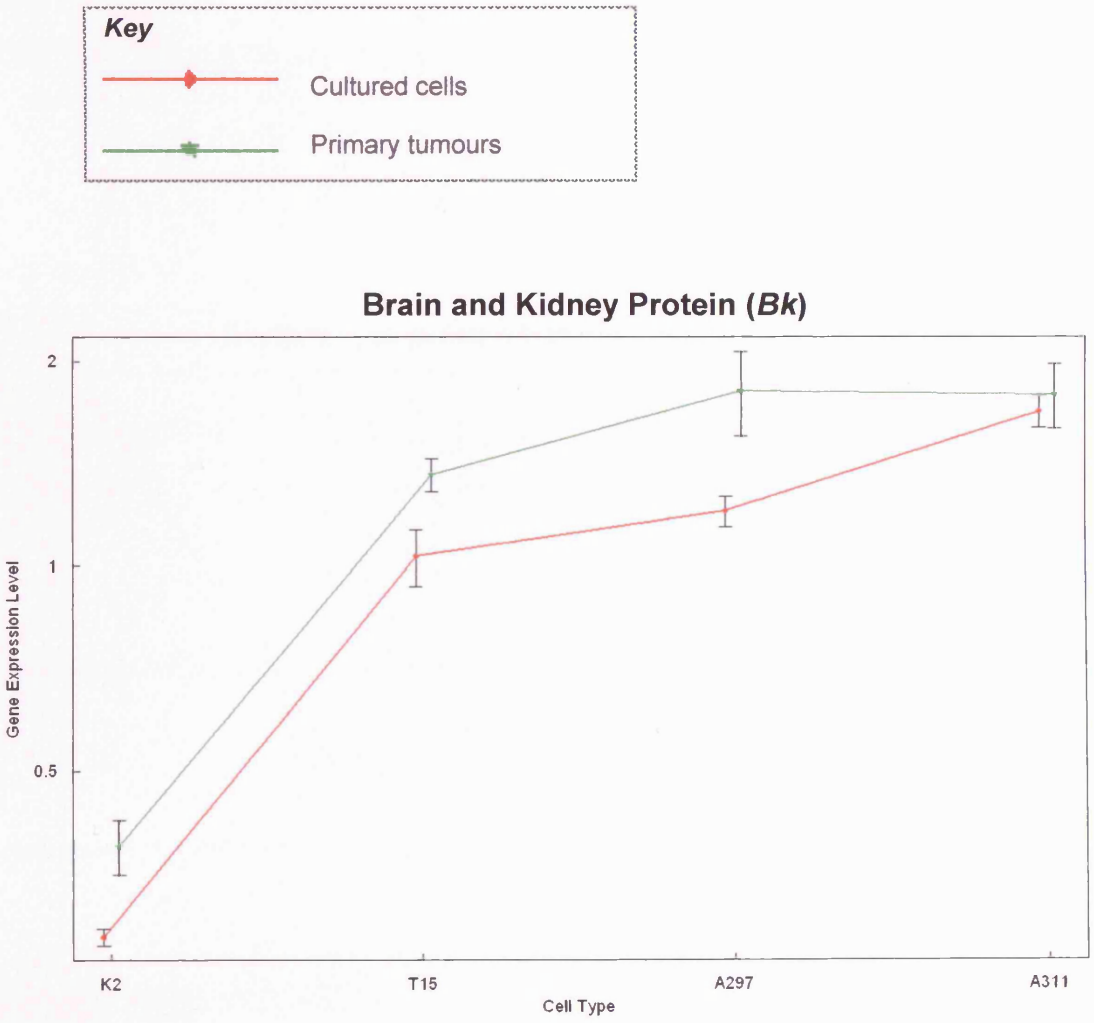
Table 6 (continued): Extended microarray experiment: Genes with greatest differential expression between non-metastatic and metastatic sarcoma cells and tumours.

Gene upregulated in metastatic cells		
Gene (gene symbol)	Gene ontology (biological process)	Expression ratio
Similar to RIKEN cDNA 9330161F08 (LOC362899)		3.0
Reversion-induced LIM gene (Ril)	Intracellular signalling; actin cytoskeleton organisation	4.5
Brain and Kidney Protein (Bk LOC497736)	Transport; cAMP-mediated signalling	4.6
EST; Affymetrix Identifier 1373642_at		5.1
EST; Affymetrix Identifier 1377065_at		5.8
Similar to Synaptopodin-2 (LOC499702)		17
Ankyrin 3, epithelial isoform g (RGD:620157)	Signal transduction; positive regulation of apoptosis	18
Chondroitin sulfate proteoglycan 4 (cspg4)	MAPK activation; cell motility; cell adhesion	65

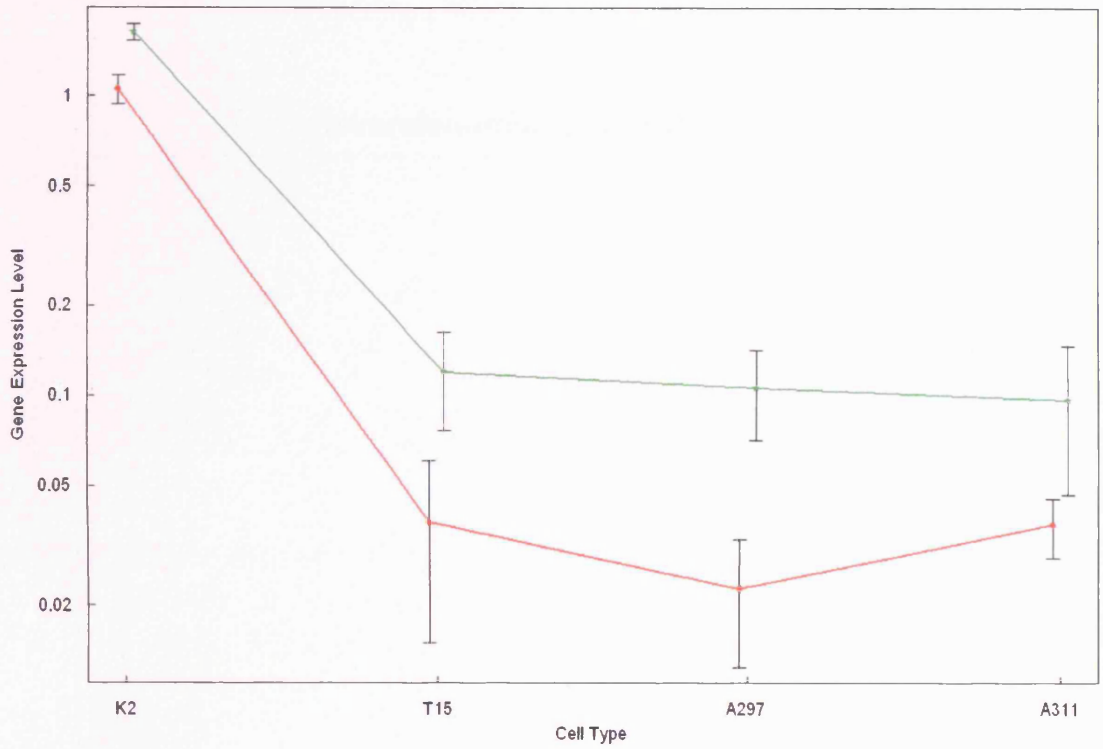
Candidate genes have an expression ratio of more than 2.5 ($P < 0.05$). GO terms for the biological process of the gene are those assigned by Affymetrix, and the first three annotations are listed here. The expression ratio indicates the mean expression value of the three metastatic cell populations compared to the non-metastatic cell population.

Figure 4.4: Expression patterns of the genes bk, cask, epb41l3 and ril.

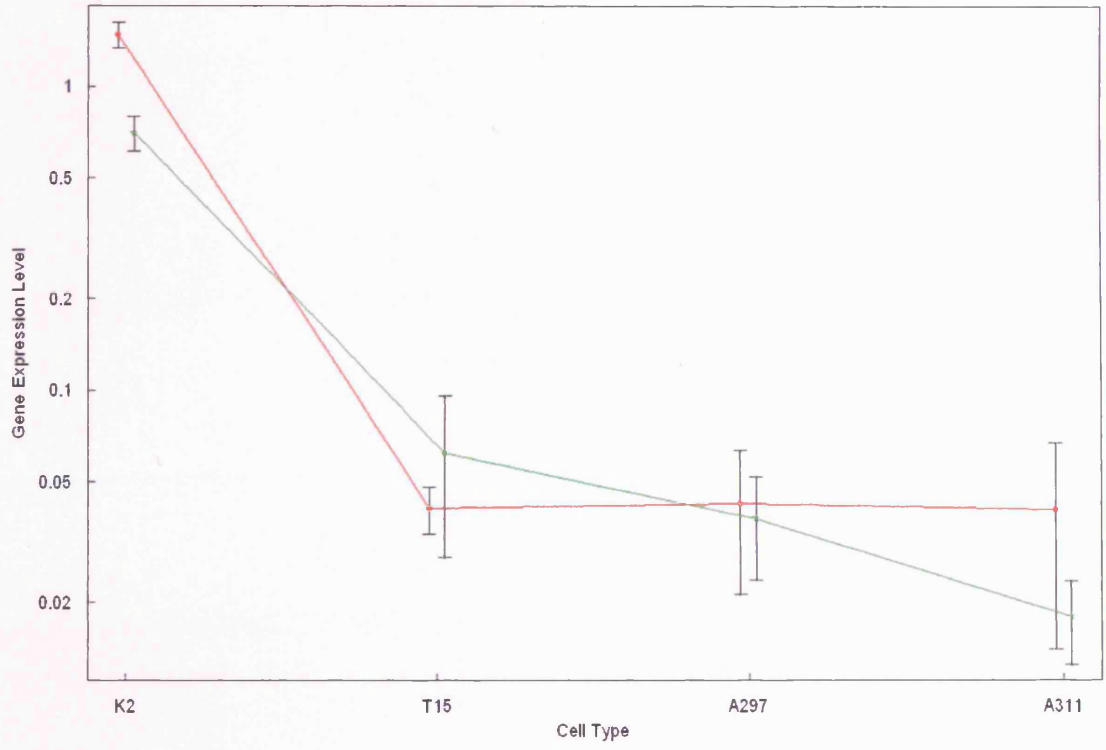
Gene expression values are shown for the cultured cells and primary tumours, where the green line represents expression levels in the cultured cells, and the red line represents expression levels in the primary tumours. In these graphs, the expression patterns of the cells and tumours are closely matched since this was one of the criteria used to define a candidate gene.

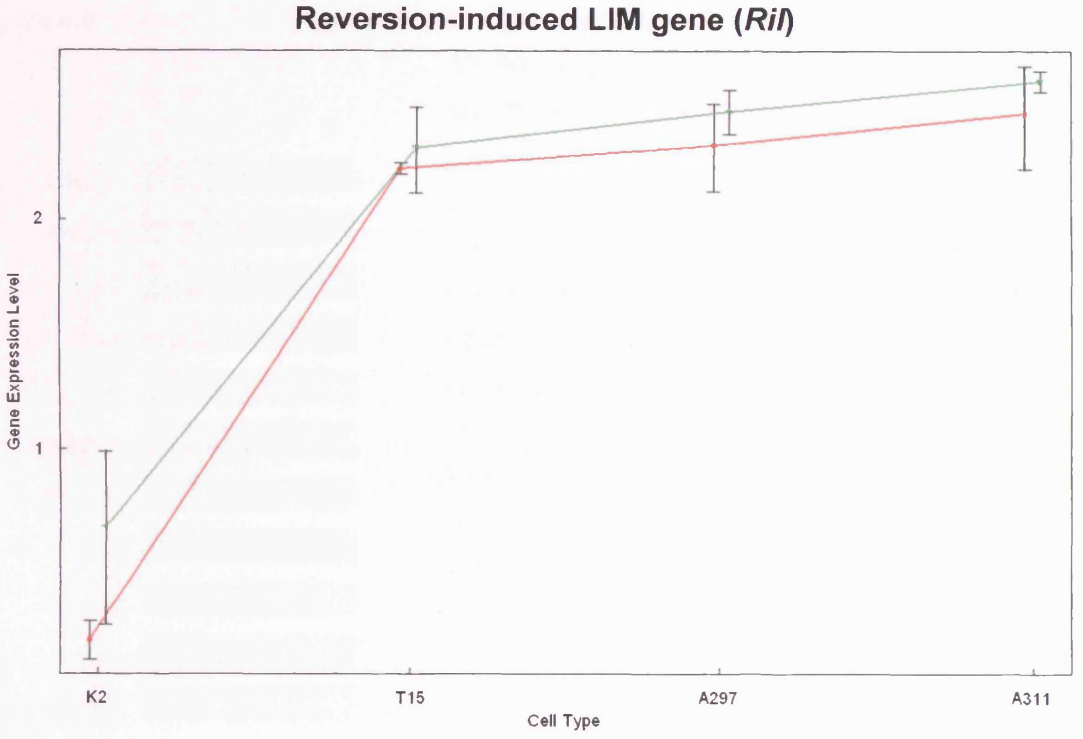


Ca/CaM dependent serine protein kinase (*cask*)



Erythrocyte protein band 4.1-like 3 (*epb41l3*)




Reversion-induced LIM gene (*Ril*)
Gene Expression Level
2
1
K2
T15
A297
A311
Cell Type

4.2.3.1 Confirmation of candidates by RT-PCR

RT-PCR for ten of the candidate genes confirmed the expression patterns found by the microarray analysis (Figure 4.5). Replication of the expression patterns for all of the genes tested confirmed that the microarray data are reliable.

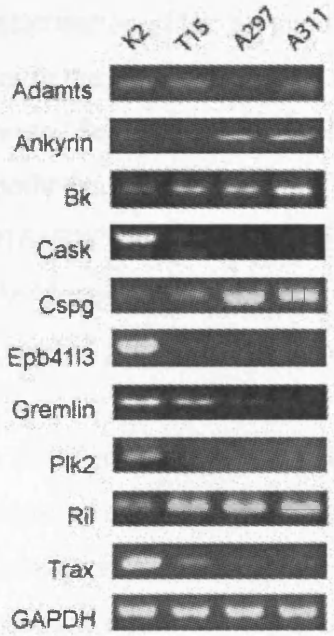


Figure 4.5: Extended microarray experiment: validation of candidate gene expression. *RT-PCR was performed for ten of the candidate genes from the extended microarray experiment. Total RNA was extracted from each cell population and reverse-transcribed into cDNA, and regions of each gene of interest were amplified by PCR. PCR products were resolved on a 1% TBE agarose gel supplemented with ethidium bromide, and visualized by UV transillumination. The expression patterns observed in RT-PCR are the same as those indicated by the microarray, confirming that the microarray data are reliable. RT-PCR for GAPDH was performed as a control.*

4.2.4 Investigation of candidate genes using the literature and gene ontology

To assess the value of pursuing each of the candidate genes further, I investigated the functions of the genes using their GO (gene ontology) terms and by searching the literature. Unfortunately, not all transcripts probed by the Affymetrix RAE230A GeneChip® array had been assigned annotation (i.e. they were unknown genes), since characterisation of the rat genome is incomplete. In fact, nine out of the twenty-three entries in the candidate gene list were not annotated (see Table 6). Five of these are Expressed Sequence Tags (ESTs). Four others were unknown, but had been named "Hypothetical gene AF152002", "Similar to CG12279-PA", "Similar to Riken cDNA 9330161F08" and "Similar to Synaptopodin-2" on the basis of their sequence homology to other genes. Due to the scarcity of information on such transcripts, the nine unannotated genes were not investigated further.

Each of the remaining 14 candidate genes was briefly investigated by a search of the literature, to establish whether it had already been associated with metastasis or behaviours related to metastasis. This yielded useful information, for example that Ankyrin 3 is a putative metastasis promoter (Bourguignon et al., 2000), and that *Epb41l3* had been identified as a tumour suppressor (Tran et al., 1999). After considering the genes in terms of their novelty, degree of characterisation, and annotation, four were selected for further study; *bk*, *cask*, *epb41l3* and *ril*.

4.2.5 Integration of microarray data with an interaction database

The construction of gene or protein networks from microarray data is important since it may enable the functional annotation of large datasets (Brazhnik et al., 2002). The microarray data described here was used in a collaborative study (Jonsson et al., 2006) to identify protein communities involved in cancer metastasis. Though the results of the collaboration did not ultimately contribute to the choice of candidate genes, the methodology that was developed represents an improvement to microarray analysis and therefore is mentioned briefly here and fully described in Appendix 5.

A collaborator, Pall Jonsson, produced a protein network (or "interactome") of putative protein-protein interactions in the rat proteome. This was achieved by performing BLAST searches of the rat genome against all proteins in the DIP and MIPS protein interaction databases (Pagel et al., 2005; Salwinski et al., 2004). The microarray data was then overlaid onto the resulting interactome, and the differentially expressed genes used as starting nodes to produce interaction networks. Further analysis was then performed on the networks, in order to identify relevant protein communities. A protein community was defined by members of the community being linked to many other members. 37 such communities were identified and are displayed in Figure 4.6. Protein communities found by this method include TGF-β and MMP communities.

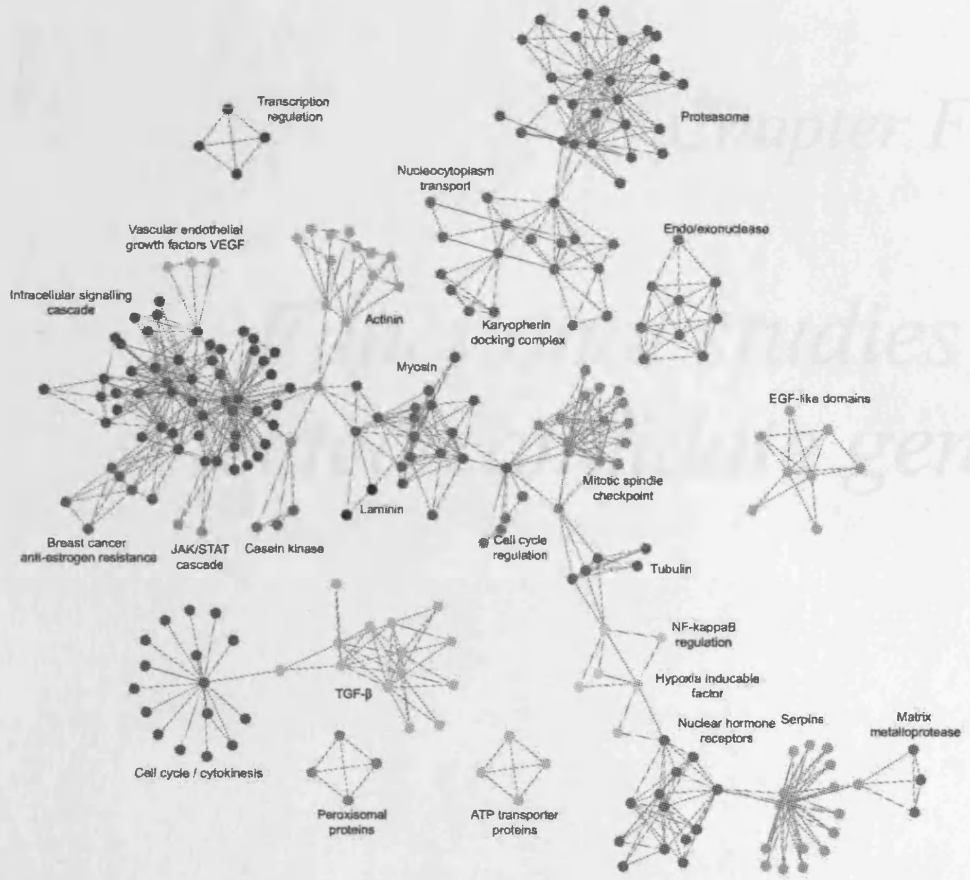


Figure 4.6: Protein communities identified by cluster analysis.

In a collaborative project with Pall Jonsson and Paul Bates, protein communities in metastasis were identified after developing a rat interactome and integrating it with the microarray data. The communities are labelled according to the main function of the proteins within the community and colours used to distinguish one community from another. This figure, prepared by Pall Jonsson, illustrates that differentially expressed genes can be clustered according to function, to reveal potentially important interactions.

Chapter Five

Functional studies of selected candidate genes

The biological functions of the candidate genes were investigated with emphasis on cell motility and the cytoskeleton. My strategy was to alter the candidate gene expression by RNAi or the expression of cDNA. RNAi was achieved by the use of sequence-specific, double-stranded RNA molecules to artificially interfere with the activity of targeted endogenous genes. For the expression of cDNA, the gene was cloned into a mammalian expression vector and then transfected into cells. The motility and F-actin cytoskeleton of the cells expressing altered levels of the gene were then assessed. HeLa cells were sometimes used in the functional studies as they are more amenable to transfection than the sarcoma cells.

This chapter describes work on α-Actinin, BK, CASK and RIL. α-Actinin was a candidate from the first microarray experiment (Table 4) and did not feature in the final gene list obtained from the extended microarray experiment. However, the functional studies on α-Actinin are included here, since they were underway before the extended microarray experiment had been started.

5.1 α-Actinin

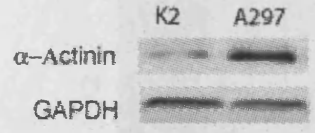
α-Actinin is an actin-binding protein that is upregulated in the metastatic A297 cells. The large pool of α-Actinin in A297 cells may enable them to perform the motility-related components of metastasis more efficiently than the K2 cells. A role for α-Actinin in the chemotaxis of A297 to PDGF/IGF is supported by the fact that it is thought to be involved in responses to IGF in other cells (Guvakova et al., 2002). α–Actinin has also been observed in invadopodia (Chen, 1989) and if the invadopodia of A297 cells are rich in α–Actinin this would be an advantage for invasion.

5.1.1 Expression of α–Actinin in the sarcoma cells

5.1.1.1 Protein expression

Immunoblotting was performed to check that there was differential expression of the protein, as well as the mRNA, between K2 and A297 cells. The immunoblot in Figure 5.1 shows that α–Actinin is overexpressed in the metastatic cells.

Figure 5.1: Expression of α-Actinin shown by immunoblotting.



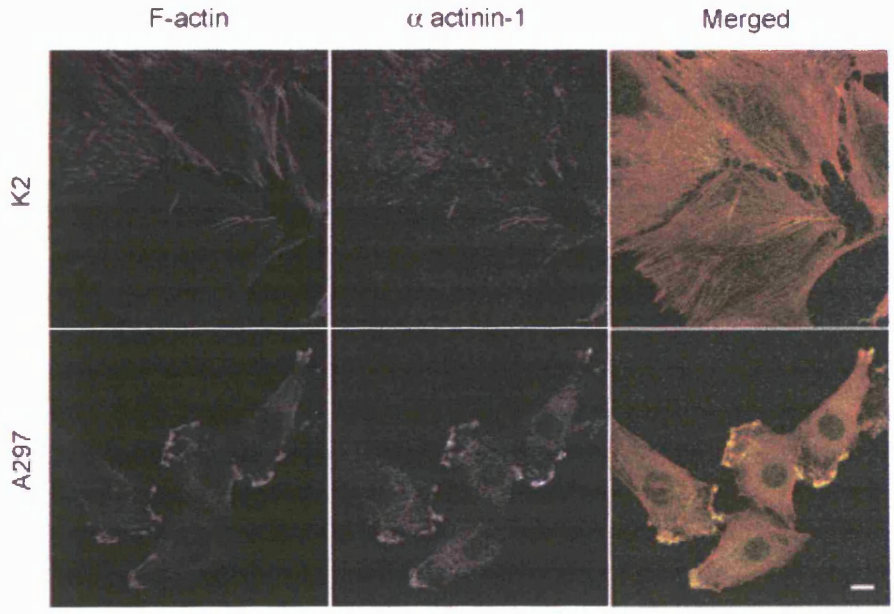
Cell lysates of K2 and A297 were produced and equal amounts of protein were resolved by SDS-PAGE, transferred onto a PVDF membrane, and probed with an antibody against α-Actinin according to standard ECL protocols. The immunoblot shows that α-Actinin is overexpressed in metastatic A297 cells. The blot was reprobed for GAPDH to confirm equal loading.

5.1.1.2 Cellular localisation of α-Actinin

The cellular localisation of α–Actinin was investigated by immunocytochemistry and confocal microscopy. Figure 5.2 shows representative images of both cell types. α-Actinin staining can be seen along stress fibres in the K2 cells, and in the ruffles of the A297 cells. The reason for the discrepancy in localisation is likely that the two cells types have their actin arranged in different ways (see Section 3.1.1), rather than that there is a cell type-dependent localization of α-Actinin to specific regions of the F-actin cytoskeleton. The reduced α-Actinin staining in K2 compared to A297 supports data from the microarray, RT-PCR and immunoblotting.

Figure 5.2: Localisation of α-Actinin in K2 and A297 cells.

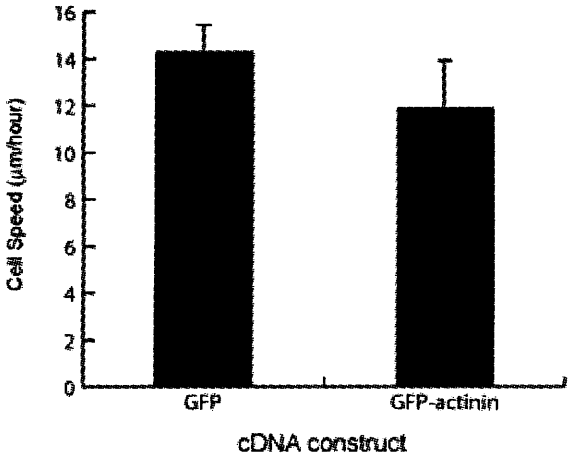
Cells were fixed and stained with rhodamine phalloidin and an antibody to α-Actinin; single confocal images are shown. α-Actinin colocalises primarily with actin in ruffles in A297 cells and actin stress fibres in K2 cells. There is slightly less α-Actinin staining in K2 than in A297 cells, as would be expected from the microarray and immunoblot data. Scale bar = 20 μm.



5.1.2 Migration of K2 cells ectopically expressing α-Actinin

There is a potential link between α-Actinin expression and cell migration, since α-Actinin is an actin binding protein. I introduced GFP-α-Actinin into the K2 cells, which do not endogenously express α-Actinin, and observed their motility. Figure 5.3 shows that control (GFP-expressing K2 cells) migrate at 14.3 ± 1 μm/hour. Since this is not significantly different from the typical K2 cell speed of 11 ± 0.5 μm/hour, it seems that microinjection and expression of GFP does not itself affect cell migration. K2 cells expressing GFP-α-Actinin migrate at 11.9 ± 2.1 μm/hour, which is not significantly different from the speed of the control cells. Therefore, the expression of α-Actinin in K2 does not increase cell speed.

Figure 5.3: Effect of α-Actinin expression on K2 cell migration.



GFP-α-Actinin or GFP alone were ectopically expressed in K2 cells and the effect on cell migration measured. K2 cells expressing GFP migrated at 14.3 ± 1 μm/hour, which is a typical K2 cell speed. α-Actinin-expressing cells migrated at 11.9 ± 2.1 μm/hour. The speeds of cells expressing GFP or α-Actinin are not significantly different. This bar chart represents data collected from 3 movies from which 24 GFP-expressing and 52 GFP-α-Actinin-expressing K2 cells were analysed. Error bars indicate the SEM.

5.1.3 Migration of A297 cells with reduced α-Actinin expression

5.1.3.1 Depletion of α-Actinin by RNA interference

To complement the overexpression experiments in the K2 cells, α-Actinin expression in A297 cells was reduced by RNAi. RNAi was achieved by introducing short interfering ribonucleotides (siRNA) into the cells by lipofection or microinjection. The immunoblot in Figure 5.4A shows that α-Actinin expression is reduced 48 hours after transfection of 100 nM siRNA into A297 cells. Figure 5.4B shows a confocal image of two α actinin-stained cells, one of which has been microinjected with a mixture of 2 μM siRNA and FITC dextran. There is no detectable staining in the FITC-positive, injected cell, confirming that the siRNA reduce α–Actinin expression. Incidentally, I later found that FITC dextran inhibits cell motility; therefore in all subsequent siRNA microinjection studies, siRNA-positive cells were identified by directly labelling the siRNA with Cy3.

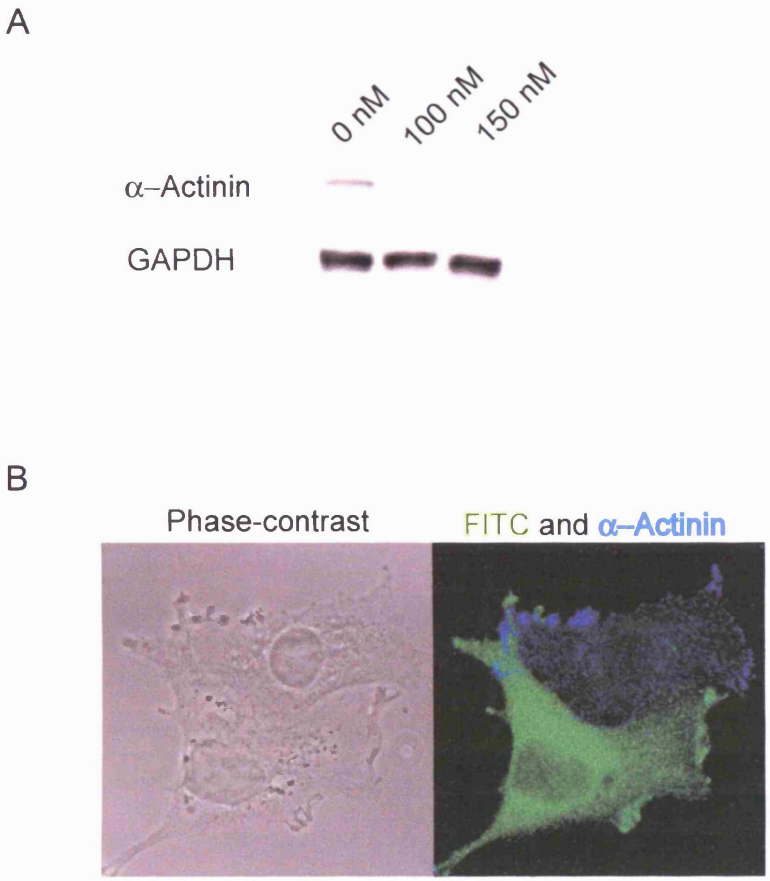
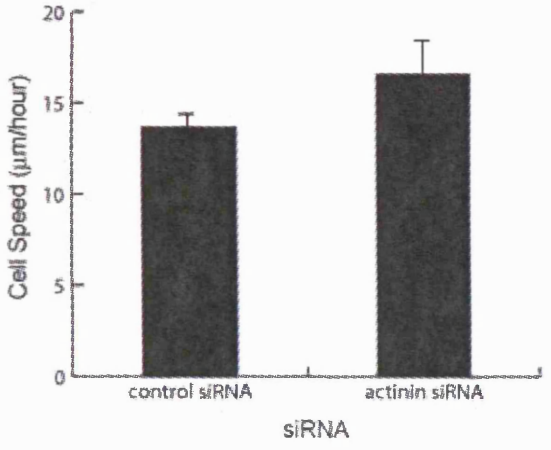


Figure 5.4: Reduction of α-Actinin expression by RNAi.
Short interfering RNA (siRNA) oligonucleotides directed against alpha actinin mRNA were transfected into A297 cells at concentrations of 0, 100 and 150 nM. After 48 hours, cells were harvested and immunoblotted. The immunoblot (A) shows that treatment with 100nM and 150nM siRNA for 48 hours is sufficient to achieve a reduction in the level of α–Actinin. For microinjection studies, siRNA mixed with FITC-dextran were microinjected at 2 μM, and the cells fixed after 48 hours and stained for α–Actinin. In (B), the FITC-negative cell has not been injected. α–Actinin can be seen in the ruffles and focal adhesions. The green cell has been injected with the mixture of FITC and siRNA. In this cell, α–Actinin staining is almost undetectable, due to the depletion of the protein by RNAi.

5.1.3.2 Migration of A297 cells with reduced α-Actinin

Figure 5.5 shows the speeds of cells microinjected with control siRNA or siRNA against *actn*. The cells migrated at speeds of 13.7 ± 0.7 μm/hour and 16.6 ± 1.9 μm/hour respectively. The speeds are not significantly different, suggesting that reduction of α actinin-1 expression by RNAi has no effect on cell migration.

Figure 5.5: Migration of A297 cells with reduced α-Actinin.



Cy3-labelled control siRNA or Cy3-labelled siRNA to actn were microinjected into A297 cells growing on coverslips. After 48 hours the coverslips were assembled into random walk chambers, and cell migration recorded by low light level digital microscopy. A297 cells microinjected with control siRNA migrated at 13.7 ± 0.7 μm/hour. Cells microinjected with siRNA to α actinin migrated at 16.6 ± 1.9 μm/hour; the speeds are not significantly different. The bar chart represents the trajectories of cells obtained from three movies, and error bars indicate the SEM.

Since this experiment and the complementary experiment of overexpression in K2 cells had no effect on migration, I concluded that α–Actinin is not responsible for the differences in cell speed. Having decided by this point to focus on genes from the extended microarray experiment, I stopped investigating α-Actinin. I did not have time to investigate its effect on cell morphology or the F-actin cytoskeleton, although preliminary data suggested that neither overexpression of α-Actinin in K2 nor its reduction in A297 had any effect on the cells.

5.2 Brain and Kidney Protein (BK)

BK protein is so-called because it was originally thought to be selectively expressed in brain and kidney. It is now known to be expressed more widely. It contains two calcium-binding motifs (C2 domains) at its C-terminus and belongs to the family of tandem C2 domain-containing proteins. Like other C2 domain-containing proteins such as synaptotagmin and rabphilin-3A, BK is tightly bound to the membrane fraction. It lacks a recognisable transmembrane region, and may target the membrane via its N-terminal cysteine cluster (Fukuda and Mikoshiba, 2001). The function of BK remains unknown, and it was therefore thought to be a potentially interesting candidate gene.

5.2.1 Expression of GFP-BK in non-metastatic K2 cells

As no antibody recognising BK was available, its protein expression in the sarcoma cells could not be confirmed by immunoblotting, and so its localisation was investigated using a GFP-BK cDNA construct. Figure 5.6 shows the localisation of GFP-BK in HeLa cells. GFP-BK is expressed throughout the cytoplasm and accumulates in the nucleus. Its wide expression gives the appearance of colocalisation, and it does not colocalise exclusively with F-actin. I would expect it to be localised to the membrane, as it is reported to be tightly bound to the membrane fraction (Fukuda and Mikoshiba, 2001). However, it is not significantly enriched at the membrane. This suggested that GFP-BK might not be correctly expressed or folded in the cells, or that the GFP tag interfered with its localisation. Therefore an untagged BK was used in subsequent studies.

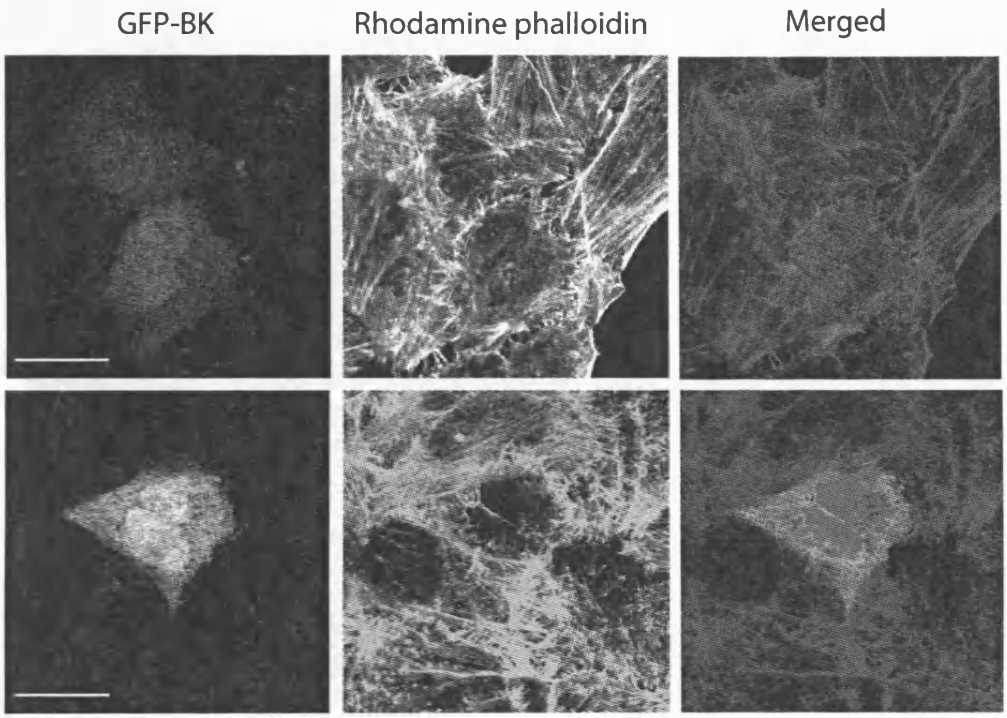


Figure 5.6: Expression of GFP-BK in HeLa cells.
Hela cells expressing GFP-BK were fixed and stained with rhodamine phalloidin. GFP-BK is expressed throughout the cell and is concentrated in the nucleus. Its ubiquitous expression gives the appearance of colocalisation with actin (see merged image in bottom panel). GFP-BK is not enriched at particular areas of the cell with the exception of the nucleus. Single confocal images are shown. Scale bar = 20 μm.

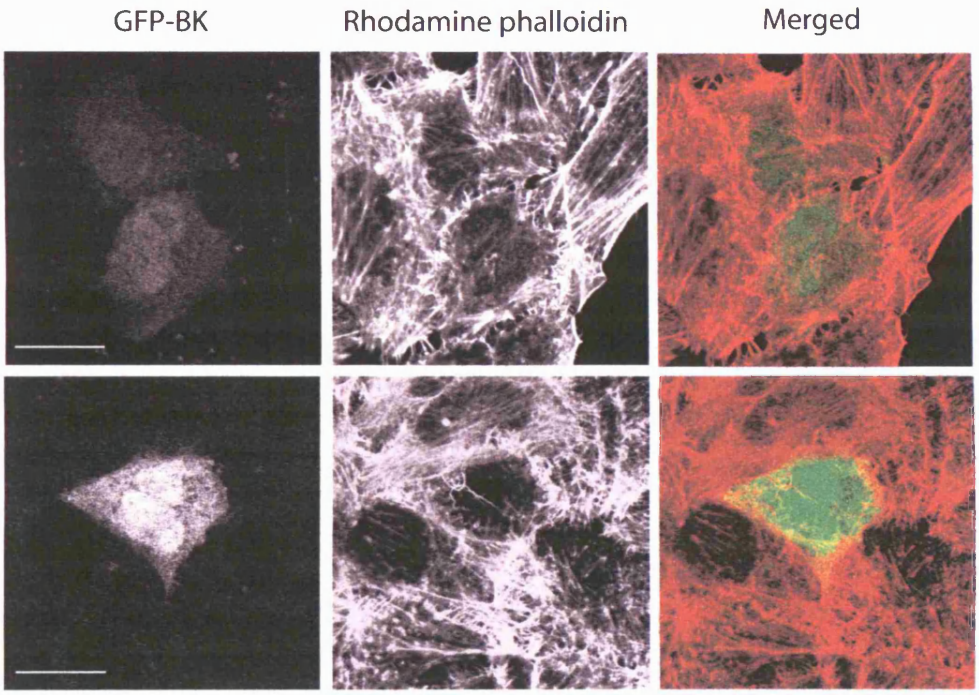
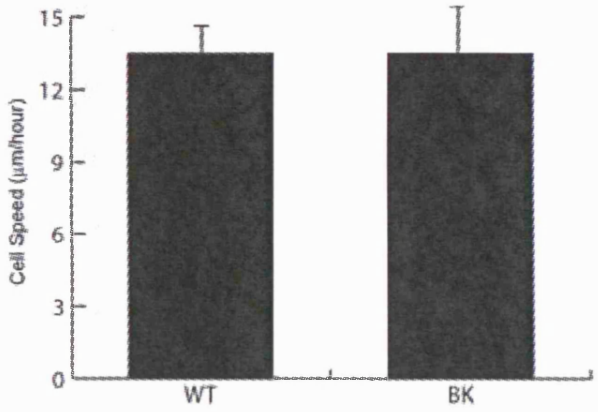


Figure 5.6: Expression of GFP-BK in HeLa cells.
Hela cells expressing GFP-BK were fixed and stained with rhodamine phalloidin. GFP-BK is expressed throughout the cell and is concentrated in the nucleus. Its ubiquitous expression gives the appearance of colocalisation with actin (see merged image in bottom panel). GFP-BK is not enriched at particular areas of the cell with the exception of the nucleus. Single confocal images are shown. Scale bar = 20 μm.

5.2.2 *Effect of BK on K2 cell motility*

The speed of K2 cells microinjected with a mixture of BK cDNA/GFP cDNA, or GFP cDNA alone, is shown in the bar chart in Figure 5.7. Both groups of cells migrated at approximately 13.5 μm/hour, indicating that the expression of BK in K2 cells has no effect on cell migration.

Figure 5.7: Effect of BK expression on K2 cell migration.



A mixture of BK cDNA/GFP cDNA, or GFP cDNA alone, was microinjected into K2 cells growing on coverslips. After 4 hours the coverslips were assembled into random walk chambers, and cell migration was recorded by low light level digital microscopy. K2 cells expressing BK and GFP migrated at 13.5 ± 1.9 μm/hour. Cells microinjected with GFP only migrated at 13.5 ± 0.9 μm/hour; the speeds are not significantly different. This bar chart represents the trajectories of 35 cells recorded in 3 movies and error bars indicate the SEM.

Some tools necessary for the investigation of BK were lacking. Firstly, no antibody was available, making it impossible to confirm the protein expression of BK in the sarcoma cells. Secondly, there was uncertainty that the cDNA construct of BK was functioning correctly; moreover, if it were functioning correctly, the experiment shown in Figure 5.7 had already proved that its expression has no effect on motility, which was one of my main interests. Therefore, it was thought prudent to stop investigating BK.

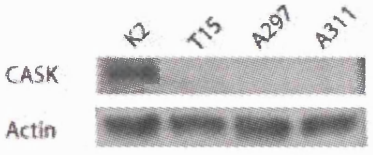
5.3 Calcium/Calmodulin Serine Kinase

Calcium/calmodulin serine kinase (CASK) is expressed in a wide range of tissues and cell lines. It shares domains with both CaM Kinases and intercellular junction proteins (Hata et al., 1996) and therefore probably transduces calcium signals to the membrane. Indeed Cohen et al. (1998) showed that it can bind to syndecan-2 and the actin-binding protein 4.1, possibly linking the extracellular matrix and the actin cytoskeleton. In this model, it is overexpressed in the non-metastatic K2 cells, and therefore represented an interesting candidate gene.

5.3.1 Expression of CASK in the sarcoma cells

Differential expression of CASK at the protein level was checked by immunoblotting (Figure 5.8). The blot shows that CASK is expressed in K2 cells, but not at detectable levels in the metastatic cells, confirming the microarray and RT-PCR data.

Figure 5.8: Expression of CASK shown by immunoblotting.



Lysates of the four cell populations were resolved by SDS-PAGE and transferred to a PVDF membrane. Blots were then probed for CASK and GAPDH according to standard ECL protocols. This blot shows that CASK is expressed in K2 cells but not in the metastatic T15, A297 and A311 cells. Actin levels indicate that equal amounts of protein were loaded.

To investigate the cellular localisation of CASK, K2 and A311 cells were fixed and stained for CASK antibody and F-actin; confocal images are shown in Figure 5.9. CASK staining is diffuse throughout the cytoplasm in both cell types, with a little colocalisation with F-actin at the lamellipodia. Confusingly, these and all other CASK immunocytochemistry images suggested that there is more CASK in A311 cells, whilst the microarray data, RT-PCR and immunoblotting indicated that levels of CASK are higher in non-metastatic K2 cells.

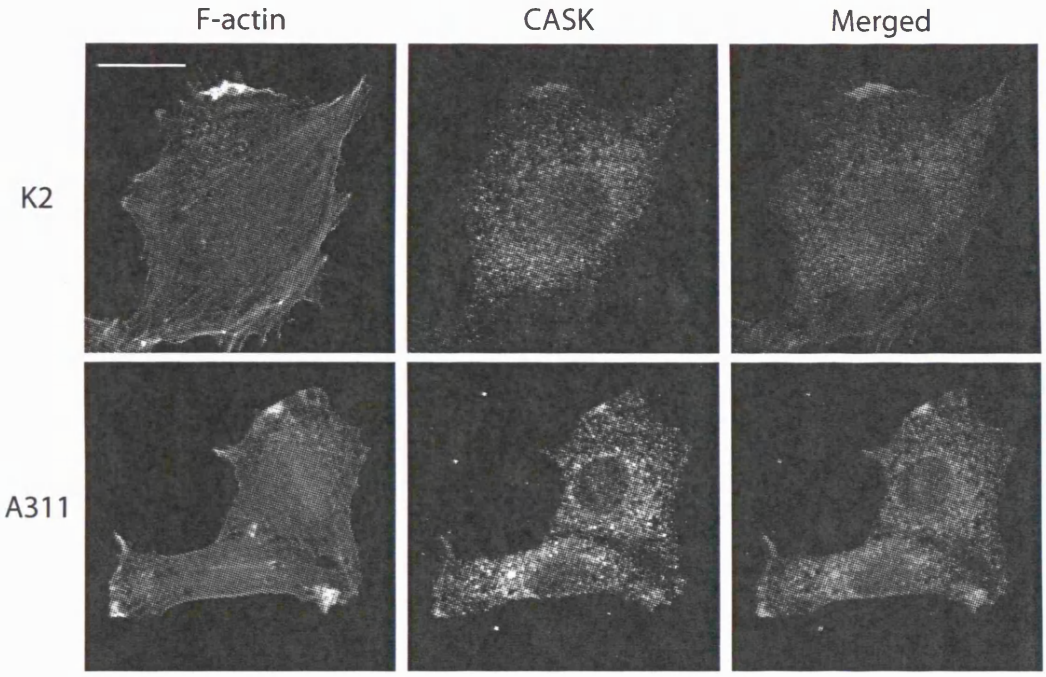


Figure 5.9: Immunocytochemistry to examine the cellular localisation of CASK. *Non-metastatic (K2) and metastatic (A311) sarcoma cells were fixed and labelled with mouse anti-CASK antibody followed by staining with rhodamine phalloidin and Alexa 488-conjugated anti-mouse secondary antibody. CASK staining is diffuse through the cytoplasm. There is a little colocalisation of CASK with ruffles in the A311 cells. The microarray data, RTPCR and immunoblotting indicate that levels of CASK are higher in K2 cells than in A311 cells. However, these and all other confocal images taken of cells stained with CASK antibody do not support this expression pattern. The images shown are 15-slice projections. Scale bar = 20 μm.*

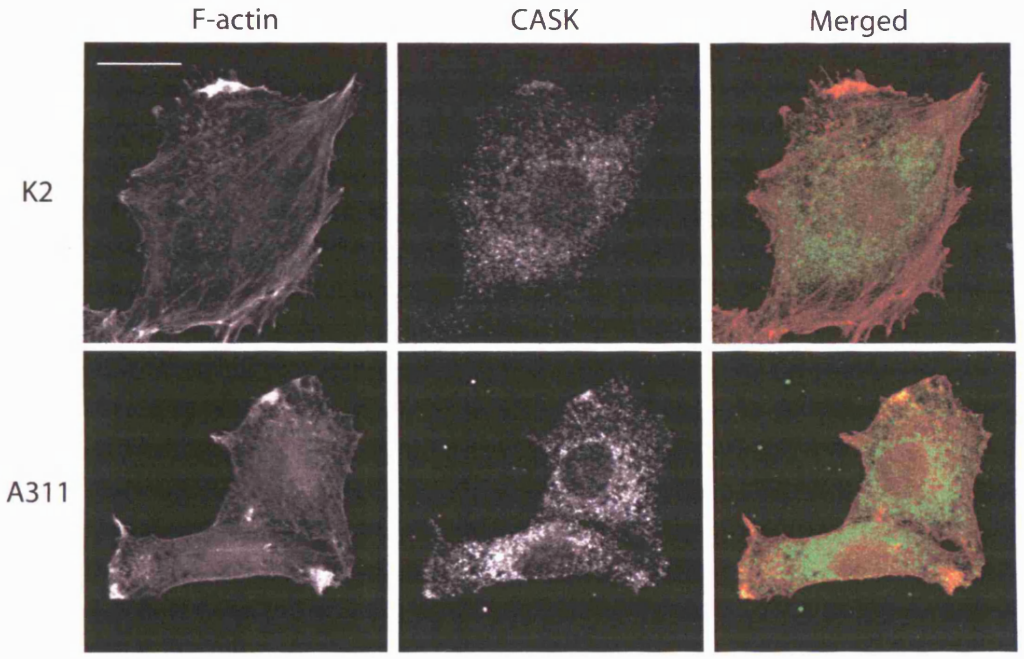


Figure 5.9: Immunocytochemistry to examine the cellular localisation of CASK. *Non-metastatic (K2) and metastatic (A311) sarcoma cells were fixed and labelled with mouse anti-CASK antibody followed by staining with rhodamine phalloidin and Alexa 488-conjugated anti-mouse secondary antibody. CASK staining is diffuse through the cytoplasm. There is a little colocalisation of CASK with ruffles in the A311 cells. The microarray data, RTPCR and immunoblotting indicate that levels of CASK are higher in K2 cells than in A311 cells. However, these and all other confocal images taken of cells stained with CASK antibody do not support this expression pattern. The images shown are 15-slice projections. Scale bar = 20 μm.*

5.3.2 Expression of GFP-CASK in metastatic A297 cells

To explore the function of CASK in the sarcoma cells, GFP-CASK was ectopically expressed in metastatic A297 cells, which do not endogenously express CASK. CASK was detected on an immunoblot of A297 cells transfected with GFP-CASK (Figure 5.10A), confirming that it was being expressed by the cells. Next, A297 cells expressing GFP-CASK were imaged on a confocal microscope. The images in Figure 5.10B show that GFP-CASK is expressed throughout the cell but is excluded from vesicles. There is no localisation to the cell membrane or with F-actin. Inspection of a number of images did not reveal any differences in cell morphology or the actin cytoskeleton in A297 cells expressing GFP-CASK.

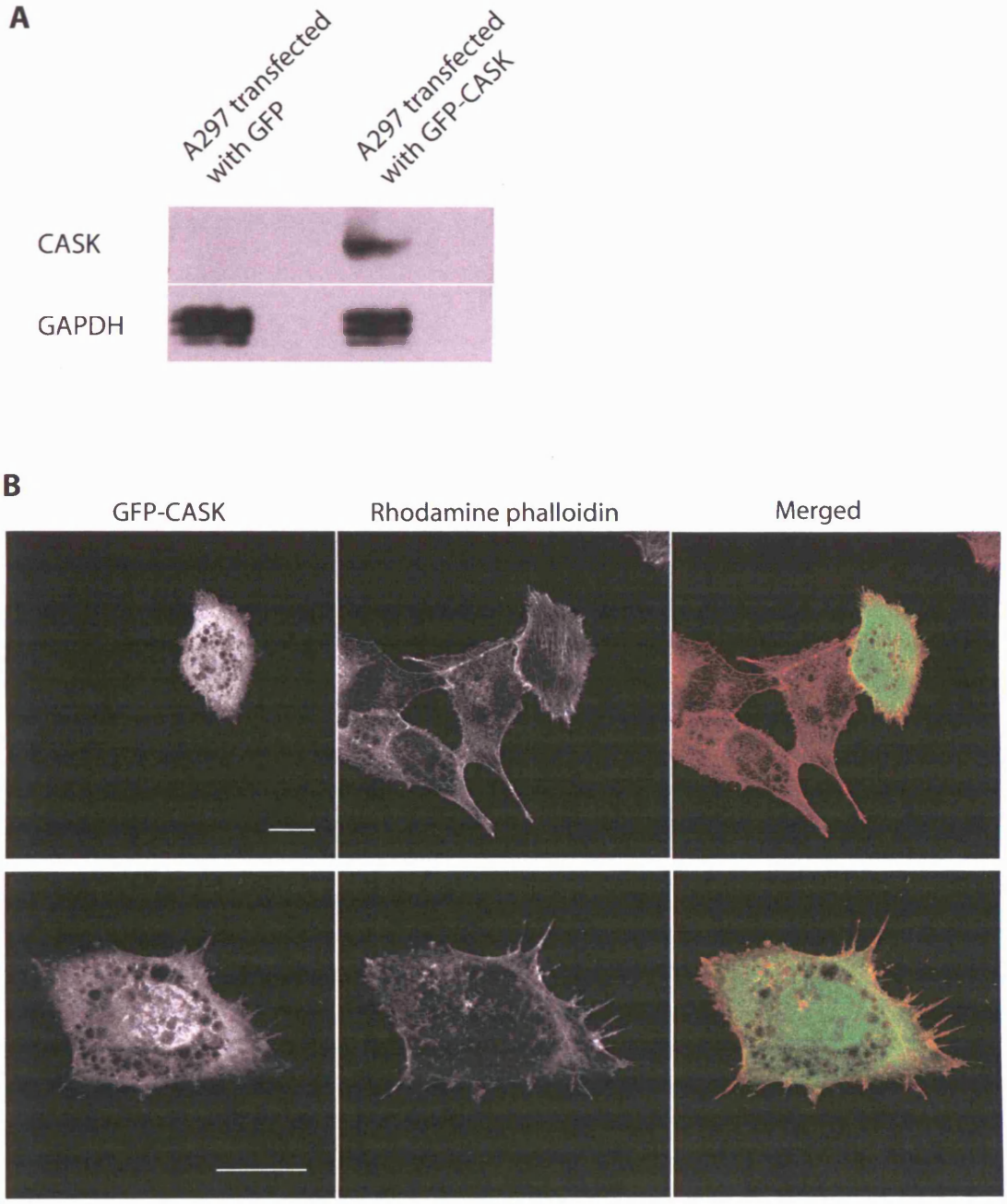


Figure 5.10: Expression of GFP-CASK in A297 cells.

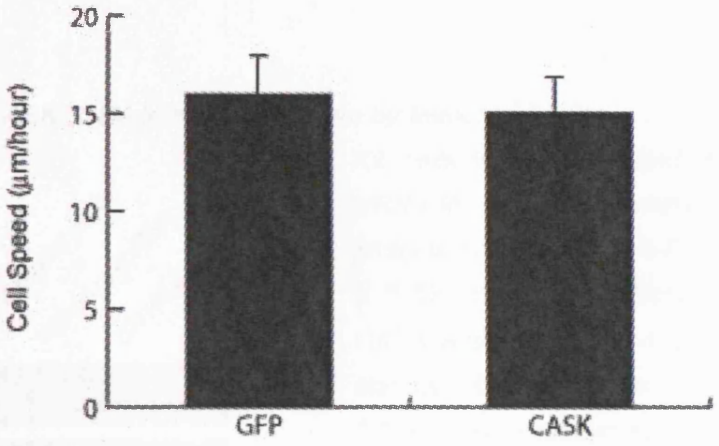
(A) Immunoblot showing CASK ectopically expressed in A297 cells. No CASK is detectable in A297 cells transfected with GFP alone. When the cells are transfected with GFP-CASK, a band of CASK can be seen.

(B) GFP-CASK was microinjected into A297 cells and the cells were fixed 4 hours later and stained with rhodamine phalloidin. Single confocal images are shown. GFP-CASK is expressed widely and is excluded from vesicles. Scale bar = 20 μm.

5.3.3 *Migration of cells overexpressing CASK*

I investigated the motility of A297 cells expressing CASK. Figure 5.11 shows that A297 cells expressing a mixture of CASK/GFP migrate at 15.1 ± 1.2 μm/hour, which is not significantly different from the speed of the GFP-expressing control cells (16 ± 1.8 μm/hr). Therefore expression of CASK in A297 cells does not affect their speed.

Figure 5.11: Effect of GFP-CASK on A297 cell migration.

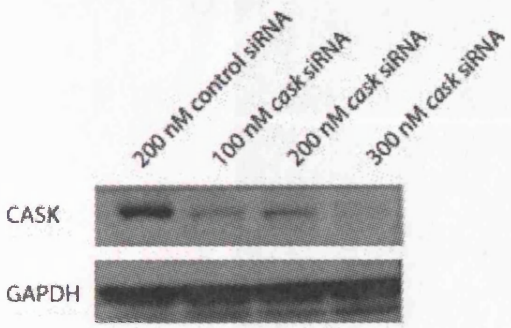


A mixture of CASK cDNA/GFP cDNA, or GFP cDNA alone, was microinjected into A297 cells growing on coverslips. After about 4 hours the coverslips were assembled into random walk chambers, and cell migration was recorded by low light level digital microscopy. Cells expressing a mixture of CASK/GFP migrated at 15.1 ± 1.2 μm/hour, whilst cells microinjected with GFP alone migrated at 16 ± 1.8 μm/hour. The speeds are not significantly different.

5.3.4 *Effect of CASK RNAi on the actin cytoskeleton*

Having looked at the effect of ectopic expression of CASK in metastatic cells, I then performed the converse experiments by reducing the expression of CASK in non-metastatic K2 cells by RNAi. To check the efficiency of the siRNA, K2 cells were chemically transfected with the siRNA at various concentrations, harvested after 48 hours, and CASK protein levels were determined by an immunoblot. Figure 5.12 shows that CASK expression is reduced 48 hours after transfection with the siRNA.

Figure 5.12 CASK RNAi in K2 cells shown by immunoblotting.



K2 cells were transfected with control or cask siRNA for 48 hours. Lysates were produced, and proteins resolved by SDS-PAGE and transferred to a PVDF membrane. Blots were then probed for CASK and GAPDH and processed according to standard ECL protocols. The blot shows that CASK is expressed in K2 cells treated with control siRNA and is reduced when the cells are treated with cask siRNA, most efficiently at 300 nM. GAPDH levels indicate that equal amounts of protein were loaded.

To study the effect of reduced CASK expression on the morphology of K2 cells, Cy3-labelled *cask* siRNA was microinjected and the cells were fixed after 48 hours and stained with rhodamine phalloidin. Cy3-positive cells were located and imaged on a confocal microscope; example images are shown in Figure 5.13. Reduction of CASK expression by RNAi does not appear to have any effect on the F-actin cytoskeleton or general morphology of K2 cells.

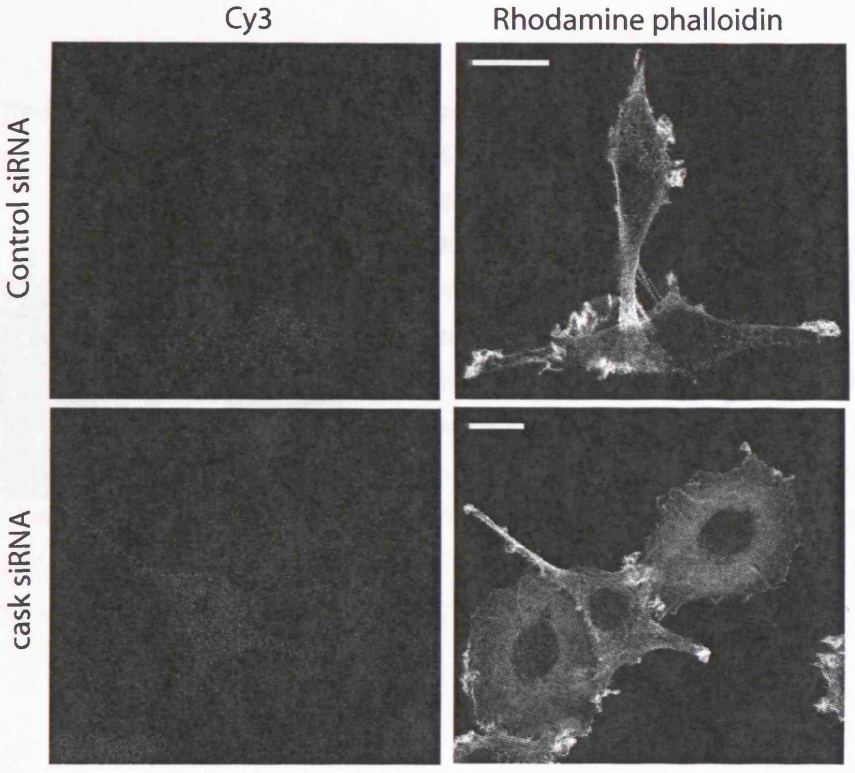


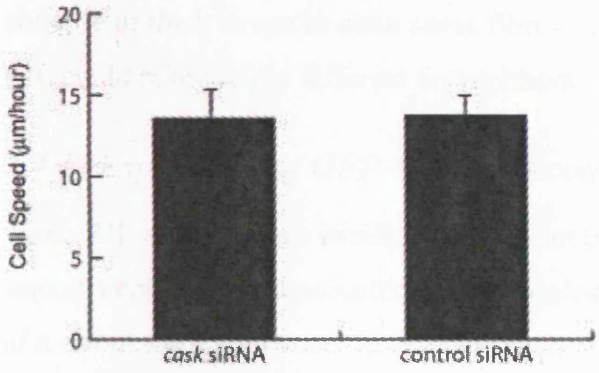
Figure 5.13: Effect of CASK RNAi on cell morphology and the actin cytoskeleton.

Cy3-labelled control siRNA or Cy3-labelled siRNA to cask were microinjected into K2 cells. After 48 hours the cells were fixed and stained with rhodamine phalloidin. Cy3-positive cells were identified and confocal images acquired. The Cy3 signals are very weak, but distinguishable from background. These are single confocal slices representative of several fields acquired from three independent experiments. Reduced CASK expression had no effect on the F-actin cytoskeleton or general cell morphology. Scale bar = 20 μm.

5.3.5 *Effect of CASK RNAi on cell migration*

Figure 5.14 shows the speeds of K2 cells microinjected with Cy3-labelled control siRNA or Cy3-labelled siRNA against *cask*. The cells migrated at speeds of 13.8 ± 1 µm/hour and 13.6 ± 1.7 µm/hour respectively. The speeds are not significantly different, meaning that the reduction of CASK expression by RNAi has no effect on K2 cell migration.

Figure 5.14: Effect of CASK RNAi on cell migration.



Cy3-labelled control siRNA or Cy3-labelled siRNA against cask were microinjected into K2 cells growing on coverslips. After 48 hours the coverslips were assembled into random walk chambers, and cell migration was recorded by low light level digital microscopy. Cells microinjected with siRNA against cask migrated at 13.6 ± 1.7µm/hour. K2 cells microinjected with control siRNA migrated at 13.8 ± 1 µm/hour. The speeds are not significantly different. This graph represents the analysis of 24 cells; error bars represent the SEM.

5.4 Reversion Induced LIM Protein

Reversion Induced LIM protein (RIL) is upregulated with metastasis in this model (Figure 4.4). It was identified as a binding partner of Protein Tyrosine Phosphatase (Cuppen et al., 1998) and is thought to be expressed in polarised epithelial cells and to play a role in the establishment and maintenance of polarity. It contains a PDZ and a LIM domain. It might act as an adaptor, recruiting molecules to the actin cytoskeleton via its PDZ domain, and binding kinases via the LIM domain. Vallenius et al. found that expression of RIL in osteosarcoma cells caused the continuous assembly and collapse of thick irregular actin stress fibres. This is interesting since the expression of RIL could relate to the different arrangements of actin in the sarcoma cells.

5.4.1 Expression of GFP-RIL in sarcoma cells

As no RIL antibody was available, it was not possible to confirm its expression in the sarcoma cells by immunoblotting. The localisation of RIL was investigated by the use of a cDNA GFP construct. In non-metastatic K2 cells, GRP-RIL localises to stress fibres (Figure 5.15A). The localisation of the protein is not due to spectral cross-talk from the rhodamine phalloidin, since GFP signal is never seen in uninjected cells (see left panel of Figure 5.15A). I looked at the effect of ectopic GFP-RIL expression on K2 and A311 cells in which it is underexpressed and overexpressed respectively. Representative images are shown in Figure 5.15B; although GFP-RIL colocalises with some stress fibres in K2 and ruffles in A311, it does not provoke any discernible alterations in the F-actin cytoskeleton.

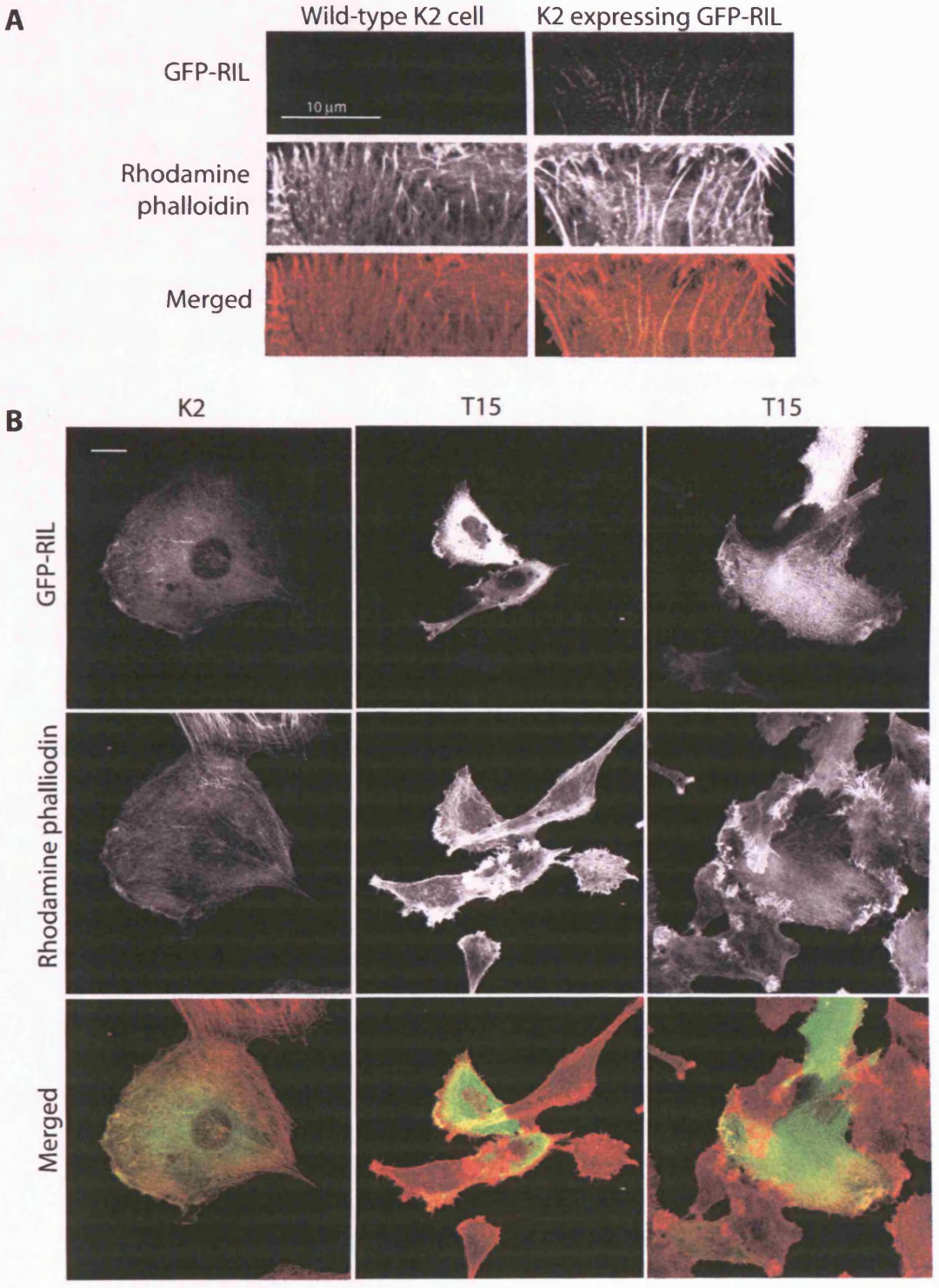


Figure 5.15: Expression of GFP-RIL in sarcoma cells.

GFP-RIL was injected into K2 cells 4 hours prior to fixing and staining of the cells with rhodamine phalloidin. Single confocal images are shown; scale bar = 20 µm.

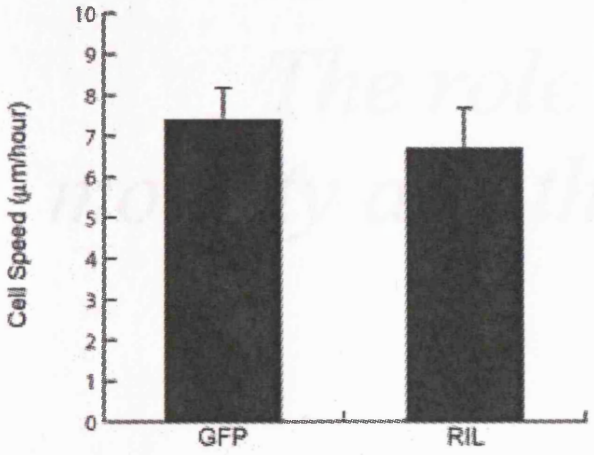
(A) Close-up of the leading edge showing that GFP-RIL localises to actin stress fibres. The left panel of images shows an uninjected cell from which no GFP signal is emitted. The images in the right panel are of a cell injected with GFP-RIL. GFP-RIL is dotted along the stress fibres.

(B) Images representing a total of three experiments investigating the effect of GFP-RIL on the F-actin cytoskeleton. The expression of GFP-RIL in K2 or T15 cells had no effect on the F-actin cytoskeleton.

5.4.2 Effect of RIL on K2 cell motility

Figure 5.16 shows the speeds of K2 cells microinjected with GFP cDNA or a mixture of RIL and GFP cDNA. Cells injected with GFP migrated at 7.48 ± 0.7 μm/hour. Cells injected with the mixture of GFP and RIL migrated at a similar speed, of 6.7 ± 0.9 μm/hour. The speeds are not significantly different, meaning that ectopic expression of RIL has no effect on K2 cell migration.

Figure 5.16: Effect of RIL expression on K2 cell migration.



RIL cDNA or GFP cDNA were microinjected into K2 cells growing on coverslips. After four hours the coverslips were assembled into random walk chambers, and cell migration recorded by low light level digital microscopy. K2 cells microinjected with GFP cDNA migrated at 7.4 ± 0.7 μm/hour. Cells microinjected with RIL cDNA migrated at 6.7 ± 0.9 μm/hour. The speeds are not significantly different.

Chapter Six

The role of 4.1B in cell motility and the cytoskeleton

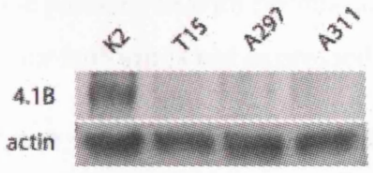
Microarray analysis showed that the gene *epb41l3* was significantly down-regulated in metastatic rat sarcoma cell lines and tumours. *Epb41l3* was originally isolated as the cDNA DAL-1 from a screen of lung adenocarcinomas (Tran et al., 1999), which was later found to be a fragment of full length *epb41l3*. Its protein product, 4.1B, has an N-terminal FERM (4.1, Ezrin, Radixin, Moesin) domain, a spectrin binding domain, a C-terminal domain, and three unique domains, U1, U2 and U3. 4.1B is thought to be a tumour suppressor, as it is lost in about 60% of lung adenocarcinomas (Tran et al., 1999) and meningiomas (Gutmann et al., 2001). I decided to focus on 4.1B since it was emerging as a tumour suppressor, but had not been related to the development of metastasis. I performed functional studies on 4.1B to investigate the potential relationship between loss of 4.1B and increase in migration, and change in actin arrangement.

6.1 Protein 4.1B is underexpressed in the metastatic cells

6.1.1 Expression of epb41l3 and its protein product 4.1B

Expression of the *epb41l3* gene, which encodes the protein 4.1B, is 36 times lower in the metastatic sarcoma cells than in the non-metastatic cells (see Figure 4.4). The expression pattern suggested by the microarray was confirmed by RT-PCR (see Figure 4.5). Immunoblotting was performed to examine the expression of protein 4.1B in the four cell populations. Figure 6.1 shows that 4.1B is expressed in non-metastatic K2 cells, and that very little is expressed in the T15, A297 and A311 cells. This indicates that the protein levels correlate with the levels of gene expression for 4.1B.

Figure 6.1: Expression of 4.1B protein shown by immunoblotting.



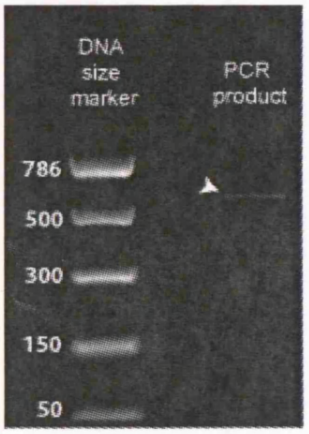
Whole cell lysates of the four cell populations were prepared and proteins resolved by SDS-PAGE and transferred to a PVDF membrane. Blots were then probed with antibodies against 4.1B and actin, and processed according to standard ECL protocols. The blot shows that 4.1B is expressed in K2 cells, and that very little 4.1B is detectable in the metastatic T15, A297 and A311 cells. Actin levels indicate that equal amounts of protein were loaded.

6.2 *Investigation of rat epb41l3 splice variants*

6.2.1 *RT-PCR to identify epb41l3 isoform*

There are two previously reported isoforms of rat *epb41l3*, the minor isoform (Genbank Accession Number AB032828), and the major isoform (NM_053927). Both isoforms can hybridise to the Affymetrix probe set for *epb41l3*. To establish which isoform of *epb41l3* is expressed in the sarcoma cells, I designed PCR primers to flank a region that diverged between the minor and major isoforms. The primers were complementary to both the major and the minor isoform, but would yield PCR products of different length; 660 base pairs for the minor isoform, and 1096 base pairs for the major isoform. RT-PCR was performed and the products were visualized on an agarose gel (Figure 6.2). A band of between 500 and 786 base pairs in length was observed. This suggested that the minor isoform was being expressed in the cells. Since no product was visible at ~1095 base pairs, even with permissive PCR conditions and high exposures, it is likely that the major isoform is not expressed in the cells.

Figure 6.2: Investigation of epb41l3 isoforms in the sarcoma cells.



Since a major and a minor isoform of epb41l3 have been reported in rat, it was necessary to establish which isoform was present in the sarcoma cells. An RT-PCR approach was used, where primers were designed to anneal to both isoforms of epb41l3, flanking a divergent region. A PCR product of approximately 660 base pairs in length suggested that the minor isoform was being expressed in the cells.

After cloning *epb41l3* from the sarcoma cells, it became apparent that the isoform in the sarcoma cells was not the minor isoform, but a third splice variant of *epb41l3*.

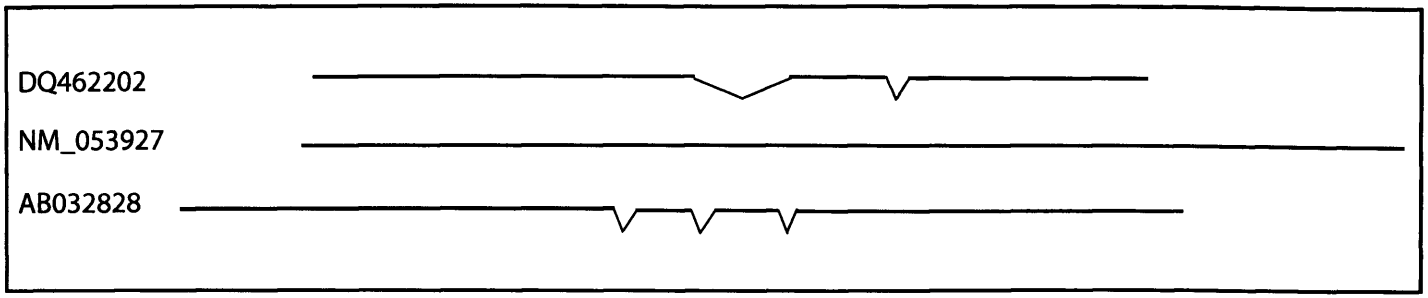


Figure 6.3: Schematic comparing the three splice variants of 4.1B.

In this figure the three variants have been aligned and regions of divergence greater than 10 base pairs in length have been indicated. The variants are largely homologous but diverge between base pairs 1500 and 2500 of the NM_053927 isoform. This region is shown in greater detail in the following figure.

6.2.2 Sequence of the new splice variant

The full length *epb41l3* gene was obtained by RT-PCR using RNA from the sarcoma cells. Once it had been cloned into the pEGFP-C3 vector, its sequence was checked and found to diverge from the sequences of the previously reported rat *epb41l3* isoforms. The sequence of the new splice variant of *epb41l3* was submitted to GenBank and assigned the Accession Number DQ462202. To investigate the common features of the three variants of rat *epb41l3*, DQ462202 was aligned with the sequence of the major isoform, NM_053927, using the Clustal W multiple sequence alignment program (Chenna et al., 2003). The sequence of DQ462202 diverges from the sequence of NM_053927 at bases 1569-2115, 2169-2204 and 2398-2521. The multiple sequence alignment covering the non-homologous regions is shown in Figure 6.4A. The chromatogram of the DQ462202 sequence covering these regions is shown in Figure 6.4B.

Figure 6.4: A new epb41l3 splice variant.

A third splice variant of epb41l3 *was identified. The sequence of the new variant, GenBank Accession Number DQ462202, was aligned with the sequence of the major isoform NM_053927 using the Clustal W multiple sequence alignment program (Chenna et al., 2003).*

(A) The multiple sequence alignment of the non-homologous regions is shown; base numbers correspond to NM_053927. The sequence of DQ462202 diverges from the sequence of NM_053927 at bases 1569-2115, 2169-2204 and 2398-2521.

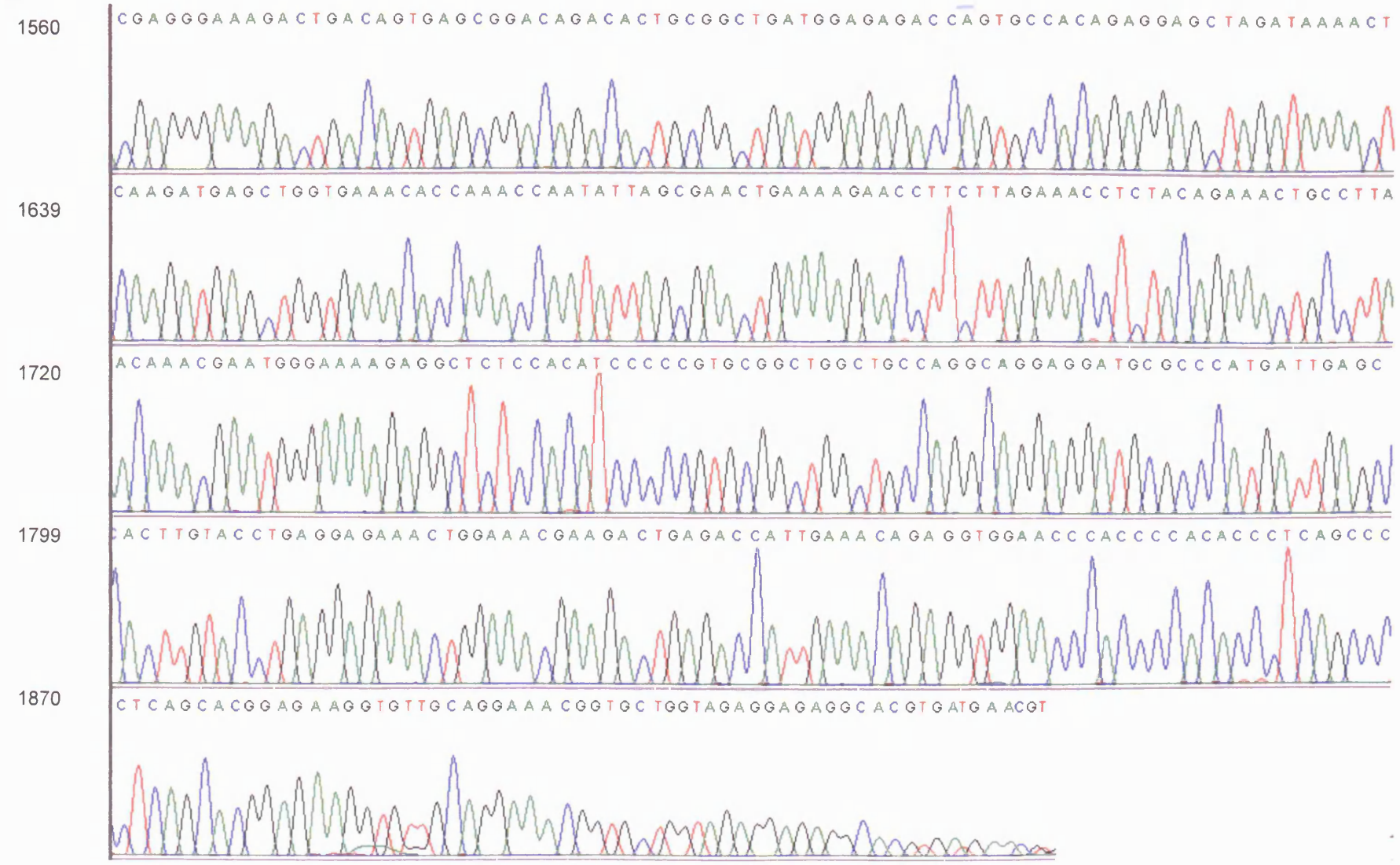
(B) The chromatogram of the DQ462202 sequence covering these regions; the sequence numbers of the chromatogram correspond to DQ462202.

A CLUSTAL W (1.83) multiple sequence alignment of novel *Epb41l3* splice variant DQ462202 and major *Epb41l3* isoform NM_053927

```
DQ462202        GAGGGAAAG---------------------------------------------------- 1569
NM_053927       GAGGGAAAGTCACCAGGGCATGGCTCTGACTCATGTCCCCTGTCACCTCCATCAGCCCAT 1620
                *********

DQ462202        ------------------------------------------------------------
NM_053927       CTTGACCCTCCCTCTCCCACAGAGCTCCGTAGGAGGTGCAAGGAGAAAGAAAGGGCCTCT 1680

DQ462202        ------------------------------------------------------------
NM_053927       CCAGGTTCTGCTGTGCCTAGAGCAGAGCTGTCCTCGGAGTCTGATGCCCAGGGGAAGGCC 1740

DQ462202        ------------------------------------------------------------
NM_053927       TACCTAGGGGACCAAGATGTGGCATTTAGTTACAGGCAGCCGGCTGGCAAAGGAACCACA 1800

DQ462202        ------------------------------------------------------------
NM_053927       CTGTTTTCCTTCTCCTTGCAGCTCCCAGAGTCATTCCCCTCGCTCCTAGATGAAGACGGG 1860

DQ462202        ------------------------------------------------------------
NM_053927       TACCTCTCTTTCCCCAACCTGTCTGAAACCAACCTCCTGCCCCAGAGTTGGCAGCATTTC 1920


DQ462202        ------------------------------------------------------------
NM_053927       TTGCCCATCCGCTCGCCCTCCCTCCTTCCCTGCTTCCTCTTCATCTTCTTCTTCCTGCTG 1980
```

```
DQ462202        ------------------------------------------------------------
NM_053927       TCTGCCTCCTTCTCAGTACCATACGCCCTCACTCTCTCCTTCCCTCTGGCTCTGTGCCTC 2040

DQ462202        ------------------------------------------------------------
NM_053927       TGCTTCCTGGAGCCCAAGGCGGCCTCCTTGAGCGCCTCCCTAGACAATGACCCGAGTGAC 2100

DQ462202        ---------------ACTGACAGTGAGCGGACAGACACTGCGGCTGATGGAGAGACCAGT 1614
NM_053927       AGTTCAGAGGAAGAGACTGACAGTGAGCGGACAGACACTGCGGCTGATGGAGAGACCAGT 2160
                      *******************************************

DQ462202        GCCACAGAG-----------------------------------GAGCTAGATAAAACT 1638
NM_053927       GCCACAGAGTCGGACCAGGAGGAAGACGCAGAGCTCAAGGCACAGGAGCTAGATAAAACT 2220
                *********                                     ***************

DQ462202        CAAGATGAGCTGGTGAAACACCAAACCAATATTAGCGAACTGAAAAGAACCTTCTTAGAA 1698
NM_053927       CAAGATGAGCTGGTGAAACACCAAACCAATATTAGCGAACTGAAAAGAACCTTCTTAGAA 2280
                ************************************************************

DQ462202        ACCTCTACAGAAACTGCCTTAACAAACGAATGGGAAAAGAGGCTCTCCACATCCCCCGTG 1758
NM_053927       ACCTCTACAGAAACTGCCTTAACAAACGAATGGGAAAAGAGGCTCTCCACATCCCCCGTG 2340
                ************************************************************

DQ462202        CGGCTGGCTGCCAGGCAGGAGGATGCGCCCATGATTGAGCCACTTGTACCTGAGGAGA-- 1816
NM_053927       CGGCTGGCTGCCAGGCAGGAGGATGCGCCCATGATTGAGCCACTTGTACCTGAGGAGACT 2400
                ************************************************************

DQ462202        ------------------------------------------------------------
NM_053927       AAACAATCTTCTGGGGAGAAGCTCATGGATGGCTCGGAGATCCTCAGTCTATTAGAGTCT 2460

DQ462202        ------------------------------------------------------------
NM_053927       GCGAGGAAACCCACAGAGTTCATAGGAGGGGTTTCATCTACTACCCAGAGCTGGGTGCAGA 2521
```

B


1560
CGAGGGAAAGACTGACAGTGAGCGGACAGACACTGCGGCTGATGGAGAGACCAGTGCCACAGAGGAGCTAGATAAAACT
1639
CAAGATGAGCTGGTGAAACACCAAACCAATATTAGCGAACTGAAAAGAACCTTCTTAGAAACCTCTACAGAAACTGCCTTA
1720
ACAAACGAATGGGAAAAGAGGCTCTCCACATCCCCCCGTGCGGCTGGCTGCCAGGCAGGAGGATGCGCCCATGATTGAGC
1799
CACTTGTACCTGAGGAGAAACTGGAAACGAAGACTGAGACCATTGAAACAGAGGTGGAACCCACCCCCACACCCTCAGCCC
1870
CTCAGCACGGAGAAGGTGTTGCAGGAAACGGTGCTGGTAGAGGAGAGGCACGTGATGAACGT

To check the domain organisation of the new splice variant, the DNA sequence of the new variant was translated into the amino acid sequence and searched for protein domains and motifs using Pfam. This showed that the FERM, SAB and CTD domains of rat 4.1B are conserved in the new splice variant (Figure 6.5). This splice variant was used in all experiments where 4.1B cDNA was expressed.

Figure 6.5: Protein domains in the DQ462202 splice variant of 4.1B.

```
  1 MTTESGSDSE SKPDQEAEPQ EAAGPQEQAG
 31 AQAAPEPAGG NVEEQPGLEQ FPEAAAHSTP
 61 VKREVSDKDR DFAAAAAKQL EYQQFEDDKL
 91 SQKSSSSKLS RSPLKIVKRP KSMQCKVTLL
121 DGSEYGCDVD KRSRGQVLFD KVCEHLNLLE
151 KDYFGLTYRD AENQKNWLDP AKEIKKQIRS
181 GAWHFSFNVK FYPPDPAQLS EDITRYYLCL
211 QLRDDIVSGR LPCSFVTLAL LGSYTVQSEL
241 GDYDPDECGN DYISEFRFAP NHTKELEDKV
271 IELHKSHRGM TPAEAEMHFL ENAKKLSMYG
301 VDLHHAKDSE GVEIMLGVCA SGLLIYRDRL
331 RINRFAWPKV LKISYKRNNF YIKIRPGEFE
361 QFESTIGFKL PNHRAAKRLW KVCVEHHTFF
391 RLLLPEAPPK KFLTLGSKFR YSGRTQAQTR
421 RASALIDRPA PYFERSSSKR YTMSRSLDGA
451 SVNENHEIYM KDSMSAAEVG TGQYATTKVI
481 SQTNLITTVT PEKKAEEEHV EEEDRRKKAE
511 EATPVAALRH EGKTDSERTD TAADGETSAT
541 EELDKTQDEL VKHQTNISEL KRTFLETSTE
571 TALTNEWEKR LSTSPVRLAA RQEDAPMIEP
601 LVPEEKLETK TETIETEVEP TPHPQPLSTE
631 KVLQETVLVE ERHVMNVHAS GDASHTARDD
661 VDATESAPAD RHSGNGKEGS SVTEAAKEQR
691 GEEADKSAPE QEQPATVSQE EDQVSAIHSS
721 EGLEQKSHFE SSTVKVESIS VGSVSPGGVK
751 LETSTKEVPV VHTETKTITY ESSQVDPGAD
781 LEPGVLMSAQ TITSETTSTT TTTHITKTVK
811 GGISETRIEK RIVITGDADI DHDQALAQAI
841 KEAKEQHPDM SVTKVVVHKE TEITPEDGED
```

The protein translation of the splice variant DNA sequence was submitted to the Pfam database and searched for protein domains. This protein sequence shows the regions of the three main 4.1B domains. The FERM domain is in red, the spectrin/actin binding domain in blue and the 4.1 family C-terminal domain in green. These three domains are conserved between the three known 4.1B isoforms, and amongst the 4.1 superfamily.

6.3 A role for 4.1B in actin organisation

6.3.1 Expression of GFP-4.1B

6.3.1.1 Cellular localisation of GFP-4.1B

As the 4.1B antibody used for immunoblotting does not work in immunocytochemistry, I examined the localisation of 4.1B in the sarcoma cells by expressing N- and C-terminally tagged GFP fusions of 4.1B cDNA. The ectopically-expressed GFP-4.1B can be expressed by the cells as it is detectable by immunoblotting (Figure 6.8A). The images in Figure 6.6 show that N-terminally GFP-tagged 4.1B is expressed in the cytoplasm of the sarcoma cells and not in the nucleus. There is some colocalisation with F-actin in lamellipodia and filopodia (indicated by the arrowhead). Expression of a C-terminally tagged 4.1B (Figure 6.7) revealed a similar pattern but with little or no colocalisation between F-actin and 4.1B.

To inspect a large number of cells, HeLa cells were tranfected with 4.1B-GFP, since they can easily be transfected with Effectene (Qiagen) and do not require labour-intensive microinjection. Images of N-terminally tagged GFP-4.1B-expressing HeLa cells are shown in Figure 6.8B. GFP-4.1B is expressed in the cytoplasm and localises to areas of apparent protrusive activity (indicated by the arrowheads). In some cells, it is possible to see membrane localisation of GFP-4.1B (Figure 6.8C).

6.3.1.2 Effect of GFP-4.1B on the actin cytoskeleton

The same images (Figure 6.6, Figure 6.7 and Figure 6.8) were examined for the possible effect of GFP-4.1B on the F-actin cytoskeleton. Inspection of the images indicated that the expression of GFP-4.1B had no effect on the cytoskeleton.

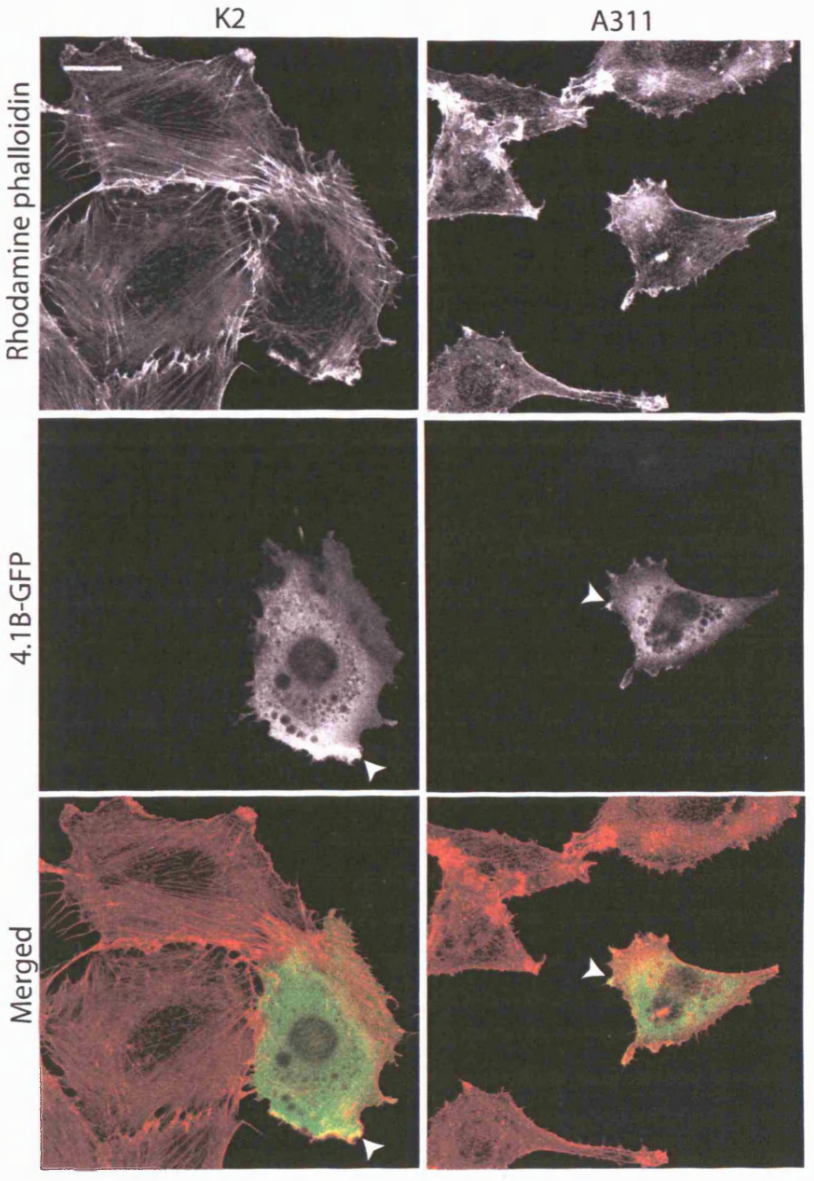


Figure 6.6: Expression of GFP-4.1B in sarcoma cells.
N-terminally tagged 4.1B-GFP was microinjected into K2 or A311 cells and the cells were fixed 4 hours later and stained with rhodamine phalloidin. 4.1B-GFP is expressed in the cytoplasm and not in the nucleus. It can colocalise with F-actin in filopodia and lamellipodia, as indicated by the arrowheads. These single confocal images are representative of three independent experiments. Scale bar = 20 μm.

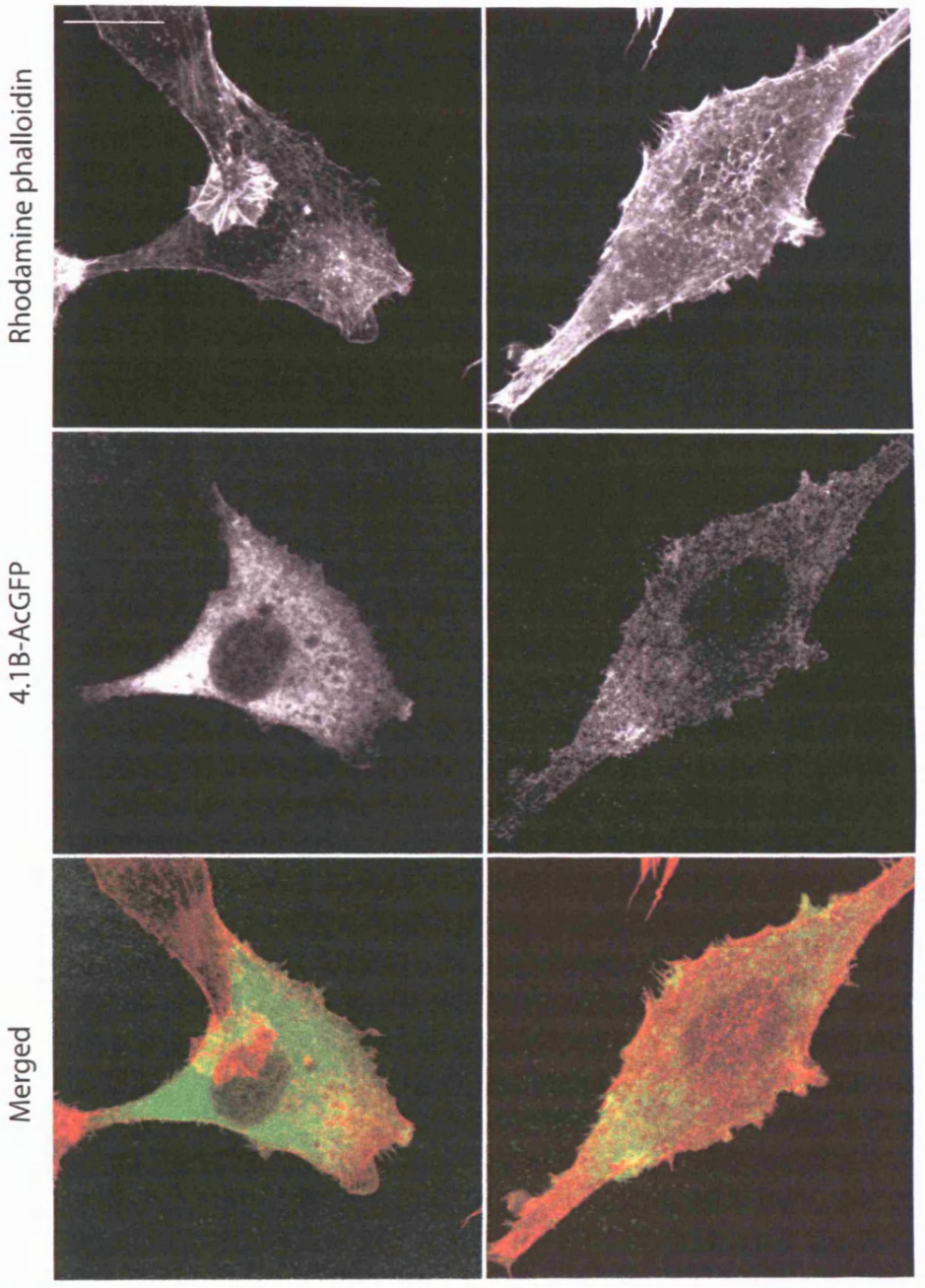
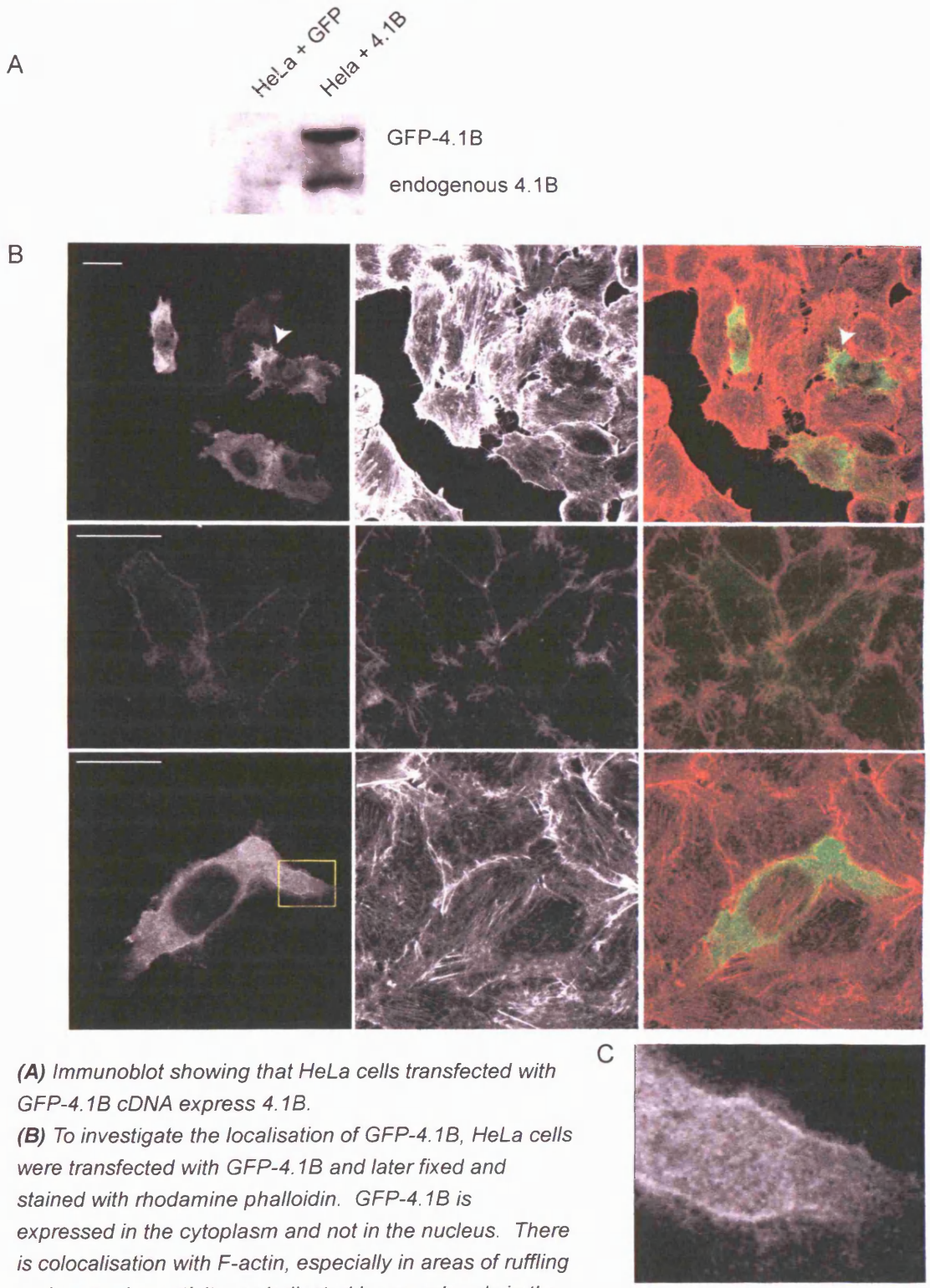


Figure 6.7: Expression of AcGFP-4.1B in sarcoma cells.
C-terminally tagged 4.1B-AcGFP was microinjected into A297 cells and the cells were fixed 4 hours later and stained with rhodamine phalloidin. 4.1B-AcGFP is expressed in the cytoplasm and not in the nucleus. There is very little colocalisation with F-actin. These single confocal images are representative of two independent experiments. Scale bar = 20 μm.

Figure 6.8: Expression of GFP-4.1B in HeLa cells.



(A) *Immunoblot showing that HeLa cells transfected with GFP-4.1B cDNA express 4.1B.*
(B) *To investigate the localisation of GFP-4.1B, HeLa cells were transfected with GFP-4.1B and later fixed and stained with rhodamine phalloidin. GFP-4.1B is expressed in the cytoplasm and not in the nucleus. There is colocalisation with F-actin, especially in areas of ruffling and protrusive activity, as indicated by arrowheads in the upper panel. There is also an enrichment of GFP-4.1B at the membrane (middle and bottom panel). This is clearer still in* **(C)** *a close-up on a region of one of the cells in B (bounded by the yellow box). These are single confocal images which are representative of three independent experiments. Scale bar = 20 μm.*

6.3.2 Effect of untagged 4.1B on the actin cytoskeleton of T15 cells

To ensure that the GFP tag did not affect the function of 4.1B, experiments were performed with an untagged cDNA. T15 cells were transfected with GFP or 4.1B cDNA, fixed with FA/PBS the next day and stained with rhodamine phalloidin. Figure 6.9 shows that there is no apparent difference between the morphology of the GFP and 4.1B-expressing cells. To be certain about the lack of effect on the actin cytoskeleton, the cells were scored according to the presence or absence of stress fibres. This analysis revealed no significant difference in actin stress fibres between cells transfected with 4.1B and those transfected with GFP (Figure 6.10).

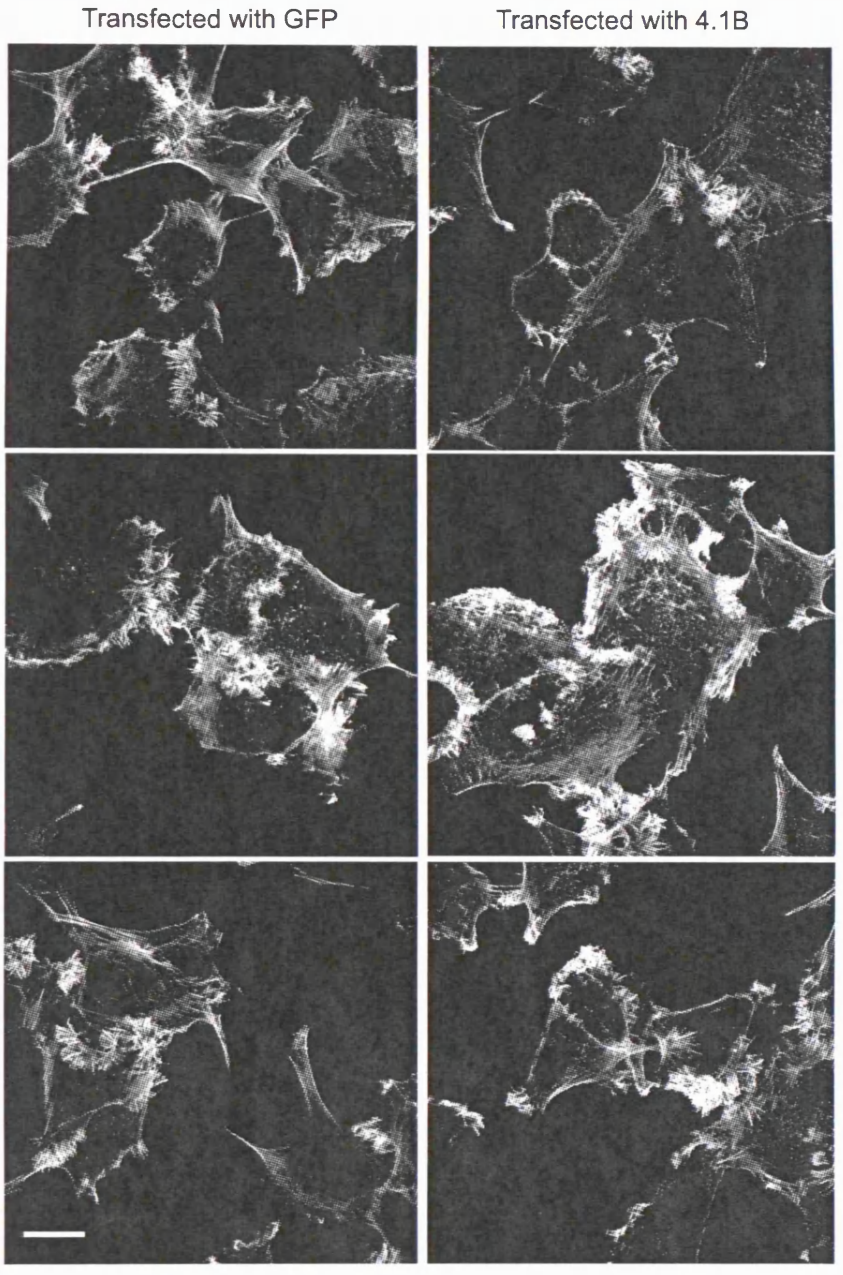
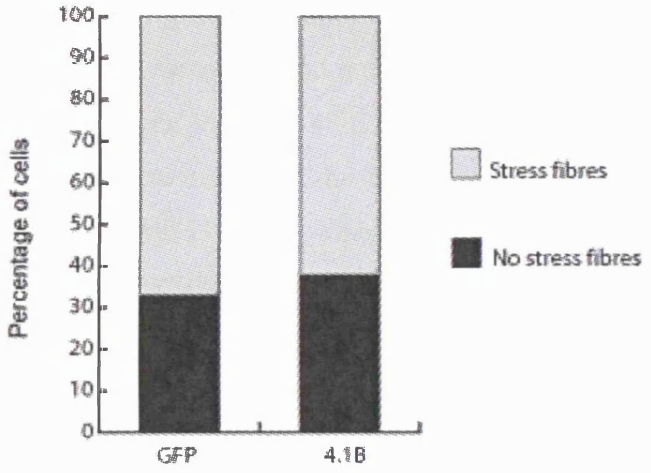


Figure 6.8: Effect of untagged 4.1B on the actin cytoskeleton of metastatic T15 cells.

T15 cells were chemically transfected with 4.1B or GFP, fixed and stained with rhodamine phalloidin. Projections of two confocal slices are shown. There is no difference between the actin cytoskeleton of the cells transfected with 4.1B and those transfected with GFP. Scale bar = 20 μm.

Figure 6.10: Quantification of the effect of 4.1B on the actin cytoskeleton of metastatic T15 cells.



The stacked bar chart shows the percentage of GFP- or 4.1B-expressing cells exhibiting at least one stress fibre, and the percentage of cells with no stress fibres. There is no significant difference in stress fibre distribution between cells transfected with 4.1B or control cells transfected with GFP alone.

6.3.3 RNAi of 4.1B disrupts the actin cytoskeleton of K2 cells

Having looked at the effect of 4.1B overexpression in metastatic cells, in which it is not endogenously expressed, I then performed the converse experiments in K2 cells which endogenously express 4.1B. I microinjected them with Cy3-labelled siRNA to *epb41l3* to reduce the expression of 4.1B, and after 48 hours, cells were fixed and stained with rhodamine phalloidin. Confocal images of the cells are shown in Figure 6.11. K2 cells injected with the siRNA to *epb41l3* display an altered F-actin cytoskeleton, with fewer stress fibres and less F-actin than the control cells.

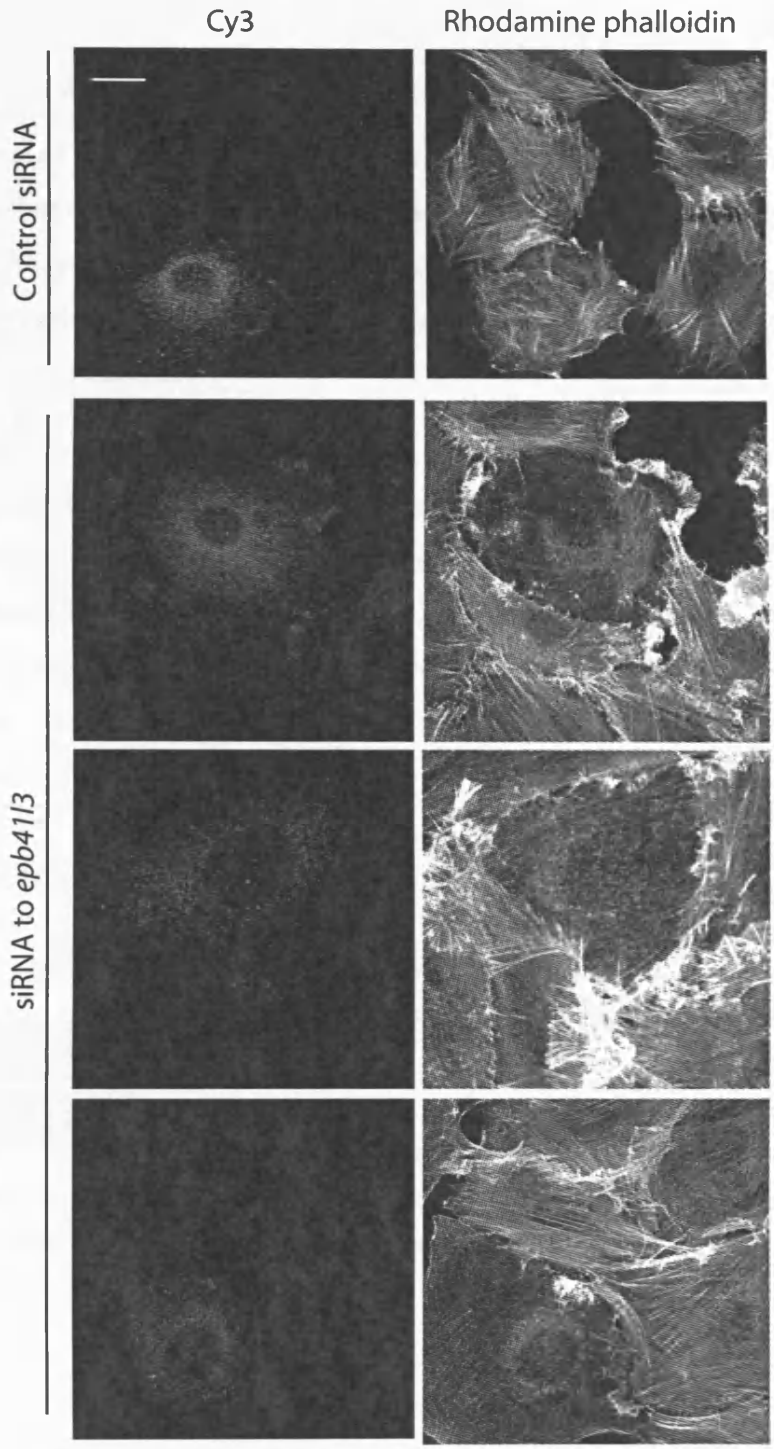


Figure 6.11: Effect of 4.1B RNAi on the F-actin cytoskeleton of non-metastatic K2 cells.

K2 cells were injected with 2 μM Cy3-labelled epb41l3 siRNA, fixed after 48 hours, stained with rhodamine phalloidin, and imaged on a confocal microscope. They display an altered actin cytoskeleton, with fewer stress fibres, and less total F-actin. Cells injected with control siRNA (top panel) are unchanged. These single confocal slices represent four independent experiments. Scale bar = 20 μm.

6.3.4 4.1B RNAi causes a loss of stress fibres in Hela cells

To quantify the effect of *epb41l3* RNAi on the actin cytoskeleton, it was necessary to analyse a larger number of cells. This would have been difficult to achieve in the sarcoma cells, since they require microinjection, which is a labour-intensive technique. Therefore, further experiments were performed in HeLa cells which, unlike the K2 cells, are highly amenable to lipofection.

6.3.4.1 RNAi of 4.1B in Hela cells

4.1B protein is expressed in HeLa cells (see lane 1 of the immunoblot in Figure 6.12). To check the efficiency of siRNA at reducing 4.1B expression, K2 cells were chemically transfected with 100 nM control or *epb41l3* siRNA, harvested after 48 hours, immunoblotted and probed with a 4.1B antibody. The immunoblot in Figure 6.12 shows that 4.1B protein expression in HeLa cells can be reduced by three independent siRNAs.

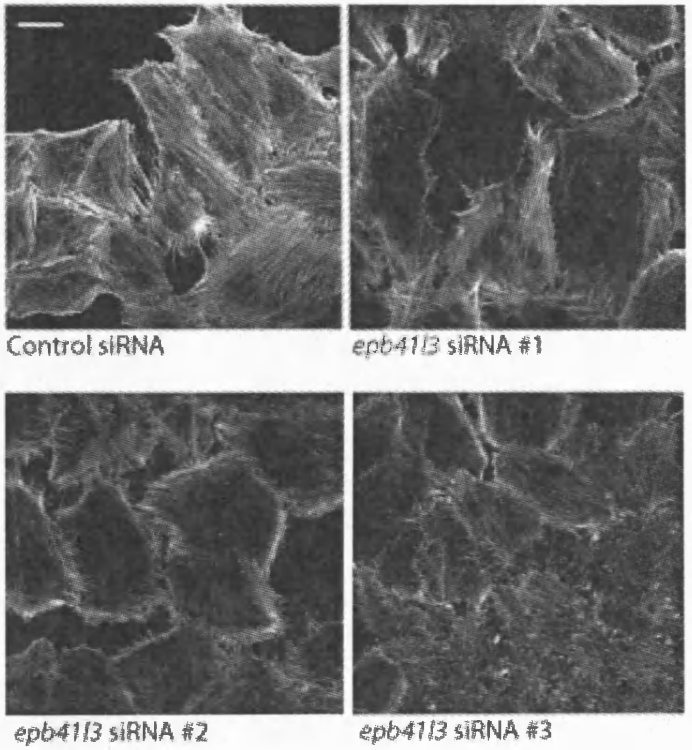
Figure 6.12: 4.1B RNAi in HeLa cells shown by immunoblotting.

HeLa cells were transfected with 100 nM control or epb41l3 siRNA and incubated for 48 hours. Lysates were resolved by SDS-PAGE and transferred to a PVDF membrane. Blots were then probed with antibodies to 4.1B and GAPDH according to standard ECL protocols. The blot shows that 4.1B is expressed in HeLa cells treated with control siRNA, and that expression is reduced when the cells are treated with three independent epb41l3 siRNAs. GAPDH levels indicate equal protein loading.

6.3.4.2 Loss of stress fibres in 4.1B RNAi HeLa cells

HeLa cells treated with siRNA to *epb41l3* for 48 hours exhibited alterations in the actin cytoskeleton (Figure 6.13). Rhodamine phalloidin staining shows that siRNA control cells often have stress fibres. However, cells with reduced 4.1B have fewer and finer stress fibres. In some cells, actin is restricted to the cortex, or arranged into short disordered filaments.

Figure 6.13: 4.1B RNAi causes a loss of stress fibres in HeLa cells.

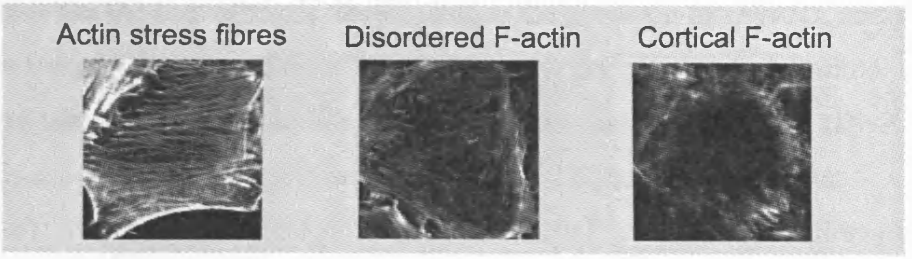


HeLa cells were transfected with control or epb41l3 siRNA, and fixed 48 hours later and stained with rhodamine phalloidin. The cells transfected with epb41l3 siRNA have fewer and finer stress fibres; actin is often restricted to the cortex, or arranged into short filaments. They also have less total F-actin than control cells. Scale bar = 20μm.

6.3.4.3 Quantification of effect of 4.1B RNAi on F-actin cytoskeleton

Confocal images of rhodamine phalloidin-stained control or 4.1B RNAi cells were subjected to a double-blind analysis, where each cell was classified according to the arrangement of its F-actin. The categories used were "actin stress fibres", "disordered F-actin" and "cortical F-actin"; examples of which are shown in Figure 6.14A. A total of 2044 cells from three independent experiments were scored, and the percentage of cells falling into each category was calculated (Figure 6.14B). Figure 6.14C shows that actin stress fibre-containing cells typically comprise just over 50% of the population when transfected with control siRNA. However, when cells are transfected with *epb41l3* siRNA, the percentage of actin stress fibre-containing cells decreases to less than 20% with an increased percentage of cells exhibiting cortical or disordered actin. The difference in distribution of stress fibres is significant (χ^2 $P < 0.005$).

A



B

	Actin stress fibres	Disordered F-actin	Cortical F-actin
Control	51.6	45.6	2.8
siRNA #1	18.5	69.5	12
siRNA #2	16.2	70.2	13.6
siRNA #3	18.5	72.2	9.3

C

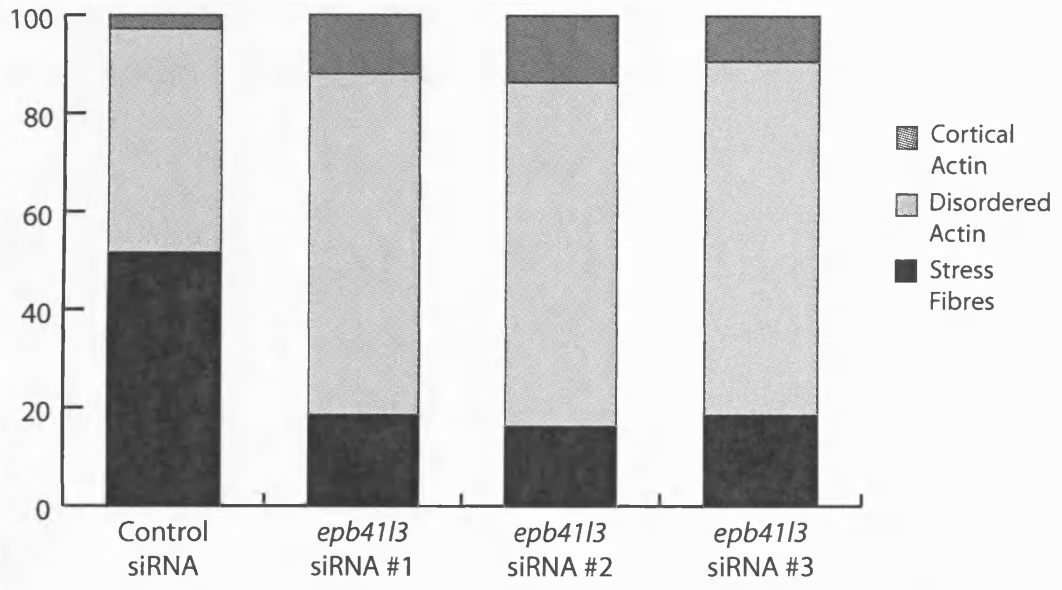


Figure 6.14: Quantification of the effect of 4.1B RNAi on the actin cytoskeleton of HeLa cells.

(A) Cells were assigned to one of three categories for quantitative analysis of the effect of epb41l3 siRNA on F-actin; "actin stress fibres", where at least one stress fibre could be seen, "disordered F-actin", where short actin filaments and puncta could be seen, and "cortical F-actin".

(B) The actin category of each cell was recorded and the percentages of cells falling into each category are displayed in the table.

(C) The stacked bar chart shows the percentage of cells in each category. Approximately half of the control cells had stress fibres, but this fell to just under 20% when the cells were treated with siRNA against epb41l3. The difference in distribution of stress fibres between control and epb41l3 siRNA cells is significant ($P < 0.005$).

6.3.4.4 Recovery of stress fibres achieved by expression of rat 4.1B cDNA

To be certain that the loss of stress fibres is due to the silencing of *epb41l3*, and not the result of an unspecific effect of the siRNA, a rescue experiment was performed. Cells were treated with *epb41l3* siRNA for 48 hours, then transfected with GFP cDNA or an RNAi-resistent 4.1B cDNA, and fixed 24 hours later (72 hours after the first transfection). The images of phalloidin-stained cells in Figure 6.15 show firstly that the 4.1B RNAi phenotype persists as long as 72 hours after the siRNA transfection. Secondly, they show that whereas the cells transfected with GFP maintain the phenotype of a disrupted F-actin cytoskeleton, the cells transfected with 4.1B cDNA recover their typical F-actin morphology. Stress fibres and organised F-actin structures have formed in cells. Quantification of the images (Figure 6.16) shows that in the RNAi-rescue cells, the percentage of cells with stress fibres is just over 40%, almost the same as that in control cells. Note that, although this may appear to be a partial rescue of the RNAi phenotype, it is likely to be a full rescue, since the transfection efficiency of the 4.1B cDNA was only between 50-70%. This rescue experiment confirmed that the effect on the F-actin cytoskeleton is due to the depletion of 4.1B by RNAi.

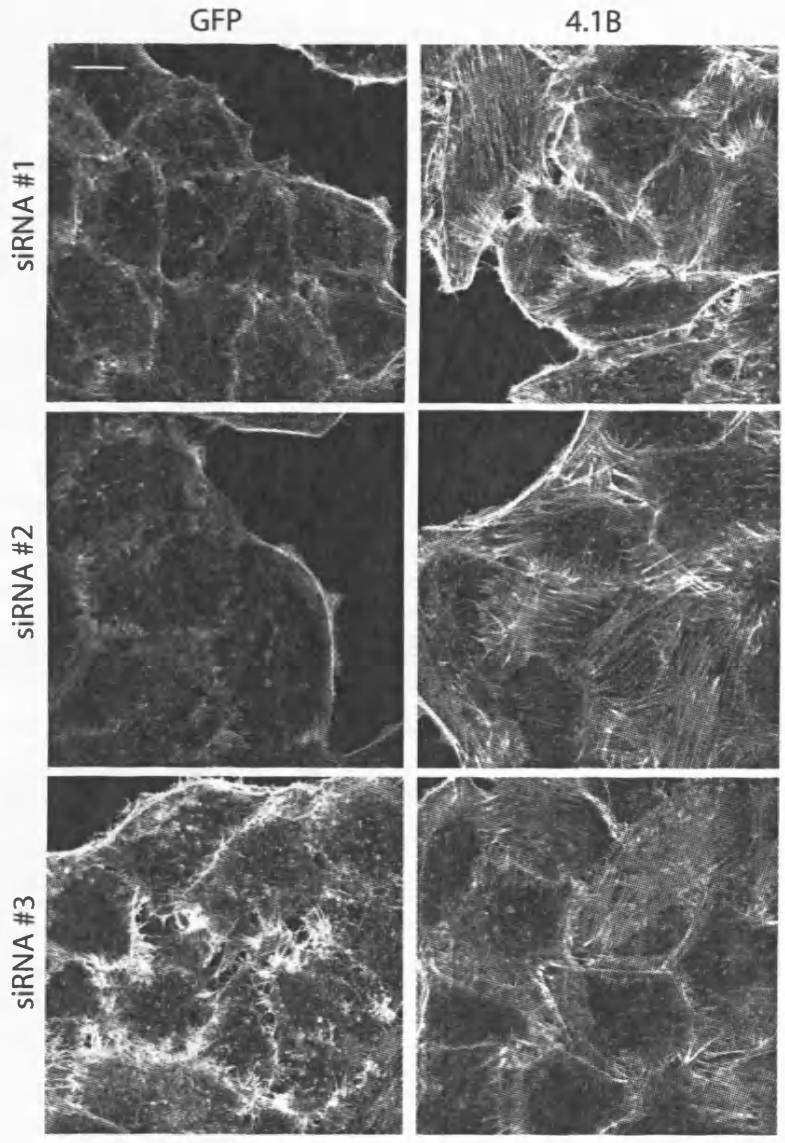


Figure 6.15: Recovery of the stress fibre phenotype in 4.1B RNAi cells by expression of an siRNA-resistent 4.1B cDNA.

HeLa cells were transfected with an siRNA-resistant 4.1B cDNA, or GFP, 48 hours after transfection with epb41l3 siRNA. After a further 24 hours they were fixed and stained with rhodamine phalloidin.

Single confocal images are shown here which are representative of three indpendent experiments. Cells transfected with GFP exhibit the 4.1B RNAi phenotype; they do not have stress fibres and their F-actin is disordered, or restricted entirely to the cortex. In contrast, cells transfected with 4.1B cDNA contain stress fibres and organised F-actin structures. This confirms that the siRNAs specifically target epb41l3. Scale bar = 20 μm.

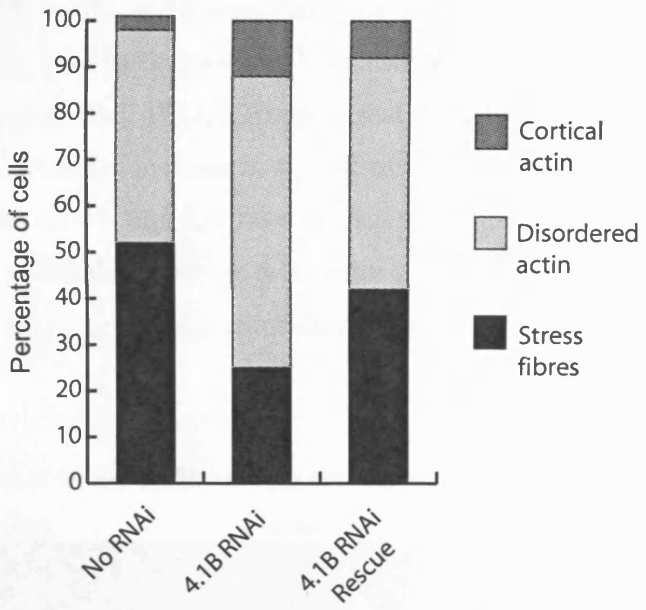


Figure 6.16: Quantification of the 4.1B RNAi rescue experiment. *The stacked bar chart shows the percentage of cells assigned to the three actin categories, which are cortical actin, disordered actin, or actin stress fibres. Of HeLa cells transfected with control siRNA, approximately 50% contained actin stress fibres. When the cells were treated with siRNA to epb41l3, this decreased to about 25% of cells. However, when the cells were transfected with an siRNA-resistent 4.1B cDNA, the percentage of cells with stress fibres increased to just over 40%. This bar chart represents data from three independent experiments, where a total of 1665 cells were scored. The difference in stress fibre distributions between the RNAi and RNAi-rescue cells is significant ($P < 0.005$).*

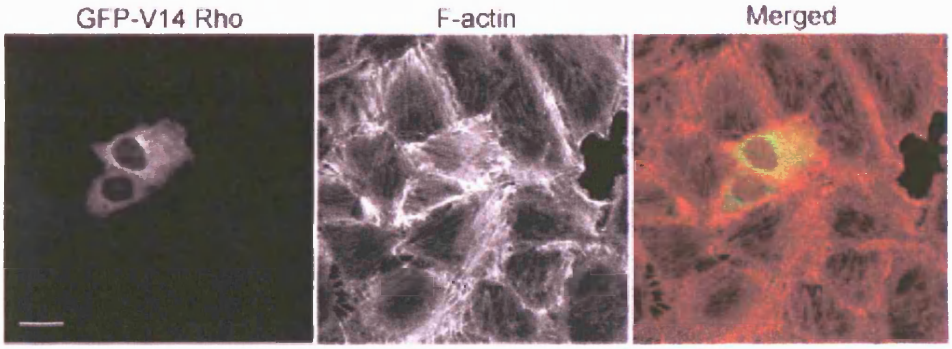
6.3.5 Abrogation of 4.1B RNAi effect by introduction of GFP-V14 Rho

I wished to explore briefly the relationship between 4.1B and Rho, a G-protein well known to be involved in the formation of actin stress fibres. I investigated whether the 4.1B RNAi phenotype of stress fibre loss could be overcome by expressing a constitutively-active mutant of Rho (V14 Rho) in 4.1B siRNA-treated HeLa cells.

6.3.5.1 Effect of V14 Rho on the actin cytoskeleton of 4.1B-RNAi Hela cells

The small G-protein Rho has been shown to induce the formation of stress fibres in fibroblasts (Ridley and Hall, 1992). To check that GFP-V14 Rho had the expected effect on the HeLa cells (an increase in actin stress fibres), the cells were transfected with GFP-V14 Rho cDNA, and fixed and stained with rhodamine phalloidin the next day. Figure 6.17 shows that there are more actin stress fibres in cells expressing GFP-V14 Rho than in the neighbouring, untransfected cells.

Figure 6.17: Effect of GFP-V14 Rho on the actin cytoskeleton of HeLa cells.



HeLa cells were chemically transfected with GFP-V14 Rho, and fixed and stained with rhodamine phalloidin. Single confocal images are shown. All cells contain actin stress fibres, but the two cells expressing GFP-V14 Rho have more stress fibres, which are also thicker (inferred from the brighter rhodamine phalloidin signal). Scale bar = 20 μm.

I next looked at the ability of V14 Rho to form stress fibres in cells depleted of 4.1B. Cells were treated with *epb41l3* siRNA for 48 hours and then transfected with GFP cDNA or GFP-V14 Rho, and fixed and stained the next day. The images of phalloidin-stained cells in Figure 6.18 show that actin stress fibres are formed in the cells transfected with GFP-V14 Rho, whereas the untransfected cells retain the typical 4.1B RNAi phenotype of a disordered F-actin cytoskeleton. To quantify the data, cells were classified according to their F-actin arrangement. The percentage of cells falling into each group is shown in the stacked bar chart in Figure 6.19. This analysis revealed that although stress fibre-containing cells account for less than 20% of the 4.1B RNAi HeLa cell population, 96% of cells transfected with V14 Rho have stress fibres. Therefore, V14 Rho can overcome the 4.1B RNAi phenotype.

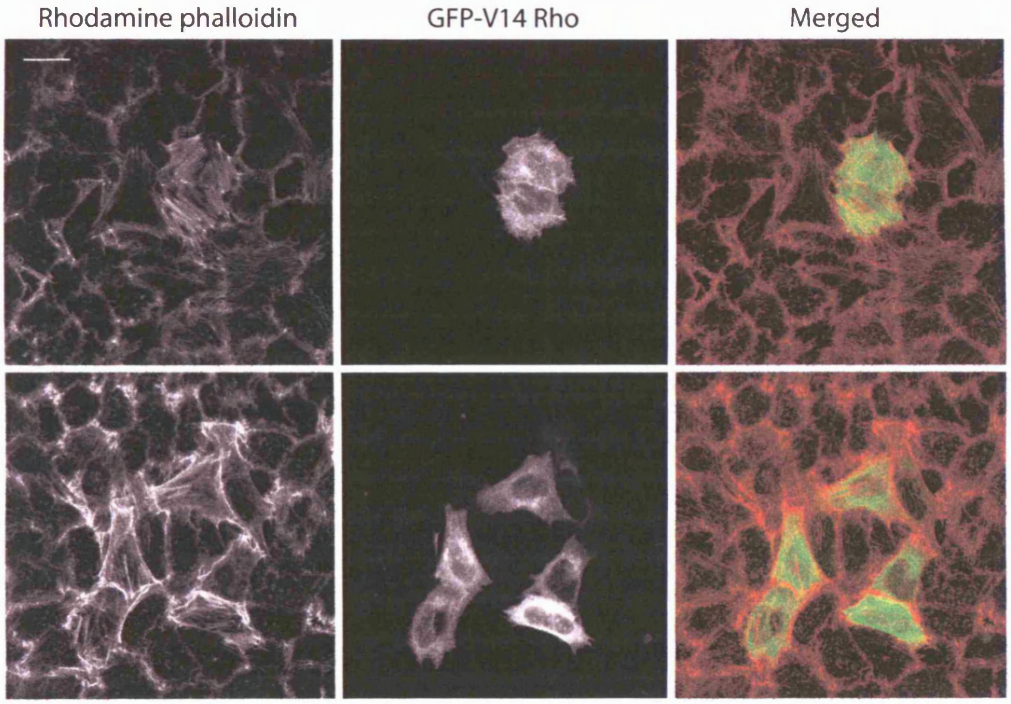


Figure 6.18: Formation of actin stress fibres in 4.1B RNAi HeLa cells expressing GFP-V14 Rho.

4.1B RNAi HeLa cells were transfected with GFP-V14 Rho cDNA, 48 hours after the first transfection with epb41l3 siRNA. Cells were fixed 24 hours later and stained with rhodamine phalloidin. These are single confocal images. Cells not expressing GFP-V14 Rho exhibit the typical 4.1B RNAi phenotype of F-actin arranged in a disordered or cortical manner. Cells expressing GFP-V14 Rho do however contain actin stress fibres. Scale bar = 20 μm.

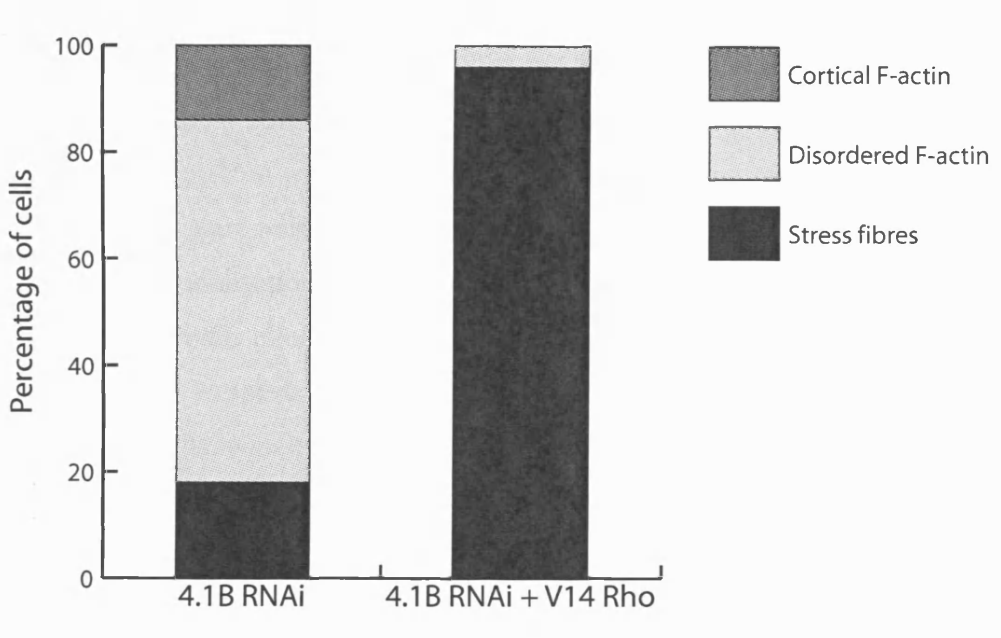


Figure 6.19: Quantification of V14 Rho-mediated stress fibre formation in 4.1B RNAi cells.

The stacked bar chart represents the percentage of GFP-V14 Rho cells exhibiting each of the following actin phenotypes; cortical actin, disordered actin, or actin stress fibres. Just under 20% of HeLa cells treated with siRNA to epb41l3 contained actin stress fibres. 96% of 4.1B RNAi cells expressing V14 Rho contain stress fibres. This bar chart represents two independent experiments, where a total of 501 GFP-V14 Rho-expressing cells were scored.

6.4 Role of 4.1B in cell motility

6.4.1 Effect of 4.1B RNAi on sarcoma cell migration

I investigated the function of 4.1B in cell motility by using siRNA to reduce 4.1B expression in the slower migrating, non-metastatic K2 cells. K2 cells growing on coverslips were microinjected with Cy3-labelled control or *epb41l3* siRNA. 48 hours later the coverslips were assembled into random walk chambers. Fields containing Cy3-positive cells were located and cell movements recorded by low light level digital microscopy. Examples of trajectories of control and *epb41l3* siRNA cells are shown in Figure 6.20A and Figure 6.20B respectively. The control cells migrated at 11 μm/hour, whilst the cells microinjected with *epb41l3* siRNA migrated at approximately 25 μm/hour. Cell speeds were normalised to the speeds of uninjected cells in the same experiment. After normalisation it was clear that the *epb41l3* siRNA cells migrated at twice the speed of the cells injected with control siRNA and this is shown in Figure 6.20C. The difference in speed is significant ($P < 0.05$). The same experiment was performed in A297 cells, which do not express 4.1B, as a control. There was no effect on cell migration (Figure 6.21). A297 cells do not express *epb41l3,* the target transcript of the siRNA, and so are unaffected by the siRNA. The lack of effect is pleasing as it confirms the specificity of the siRNA.

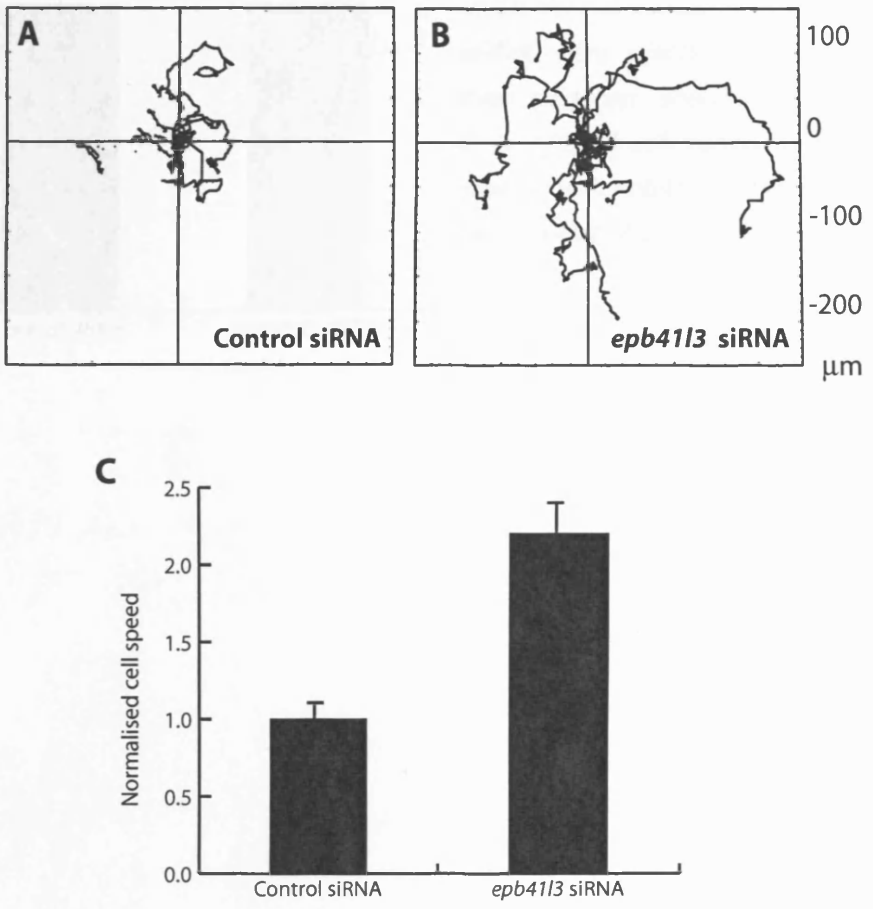
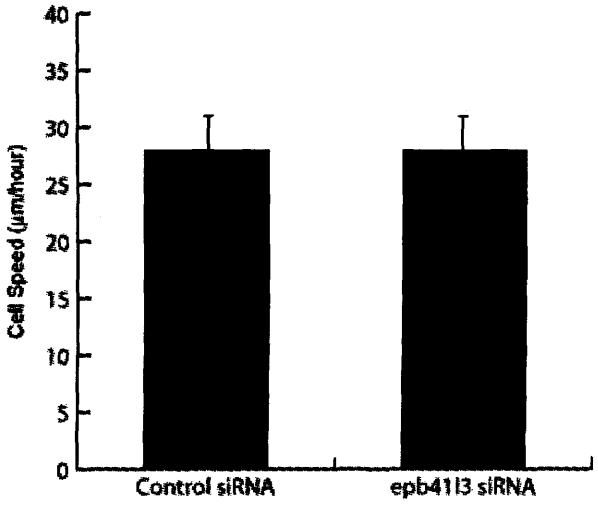


Figure 6.20: Effect of 4.1B RNAi on non-metastatic K2 cell motility.
Cy3-labelled control siRNA or Cy3-labelled siRNA to epb41l3 were microinjected into K2 cells growing on coverslips. After 48 hours the coverslips were assembled into random walk chambers, and cell migration recorded by low light level digital microscopy. Cell translocations were interactively tracked, and speeds evaluated in Mathematica. The cell trajectories represent one movie of cells injected with control siRNA (A) and one movie of cells injected with epb41l3 siRNA (B). K2 cells microinjected with control siRNA migrated at approximately 11 µm/hour. However, cells microinjected with siRNA to epb41l3 migrated at speeds of up to 25 µm/hour. Cell speed doubles when 4.1B is depleted by siRNA treatment (C) and the difference in speed is significant ($P < 0.05$).

Figure 6.21: Effect of 4.1B RNAi on metastatic A297 cells.

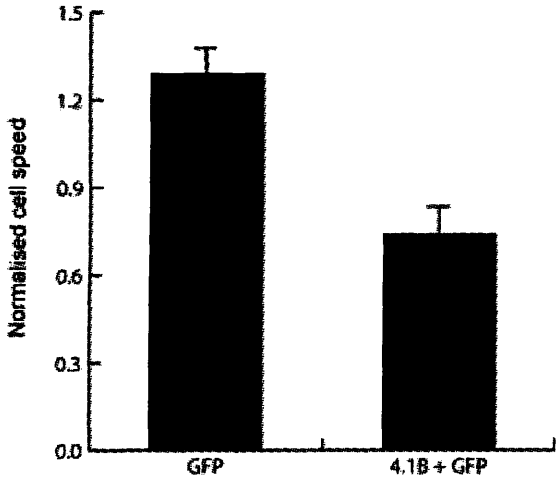


To exclude the possibility that the effect on migration of 4.1B RNAi was due to an unspecific effect of the siRNA, A297 cells, which do not contain the target of ep41l3 siRNA, were injected with the siRNA and their migration analysed 48 hours later. Cells injected with control siRNA and those injected with epb41l3 siRNA migrated at the same speed of 28 µm/hour.

6.4.2 Effect of 4.1B expression on sarcoma cell migration

Having seen an increase in cell speed in 4.1B-depleted K2 cells, I performed the complementary experiment by overexpressing 4.1B cDNA in T15 cells, which do not endogenously express detectable amounts of 4.1B protein. Cells were microinjected with GFP alone or a mixture of GFP/ 4.1B cDNA and their movements recorded by low light level digital microscopy. Cell movements were interactively tracked and analysis revealed differences in speed between cells expressing 4.1B/GFP and cells expressing GFP alone. Figure 6.22 shows that the expression of 4.1B in T15 cells causes a reduction in cell motility; cells injected with GFP migrated at approximately twice the speed of those injected with 4.1B/GFP. T15 cells microinjected with only GFP migrated at a speed not significantly different to uninjected cells.

Figure 6.22: Effect of 4.1B expression on metastatic T15 cell motility.



GFP cDNA, or a mixture of GFP and 4.1B cDNA, was microinjected into T15 cells growing on coverslips. After five hours the coverslips were assembled into random walk chambers, and cell migration recorded by low light level digital microscopy. Cell translocations were interactively tracked, and speeds evaluated in Mathematica®. Cells microinjected with both GFP and 4.1B migrated more slowly, and cell speed was reduced to approximately half that of the GFP-injected cells. The difference is significant ($P < 0.05$).

Chapter Seven

Discussion

The aim of this project was to improve our understanding of the gene expression underlying metastasis by conducting a gene expression microarray analysis of a rat sarcoma model of metastasis. The results, approaches, and challenges encountered during my research are discussed below, together with the experiments that time did not allow, and future directions this research could take.

7.1 The rat sarcoma model of metastasis

7.1.1 Suitability of the rat sarcoma model for microarray analysis

Microarray analysis is now a relatively commonly used approach to investigate metastasis. This thesis attempts to deliver an incremental improvement to this technique, principally through the use of a panel of four cell lines which originate from a common, spontaneous transformation, and which have different abilities to cause metastases. Previous microarray studies of metastasis have relied upon samples from cancer patients, or animal models with genetic heterogeneity between samples and between subjects. The model of metastasis used in this thesis was developed in inbred rats, thus providing a uniform genetic background for the experiments. This helped to reduce the noise in the microarray data and increased the likelihood that the genetic differences that were observed were due to the different abilities of the cells to produce metastases.

The rat sarcoma model was suitable for microarray analysis for other reasons. Most importantly, the cells exhibited differences in metastatic potential. Secondly, the model does not use mice which lack an immune system, such as nude mice, thus allowing the effects of the immune system to be taken into account during metastasis assays. This is important since the immune system is known to play a role in metastasis, for example via the activity of tumour associated macrophages. Finally, the sarcoma cells exhibit *in vitro* behaviours which could be related to their *in vivo* metastatic potential, and these offered opportunities for assaying the effects of genes revealed by the microarray.

7.1.2 In vitro behaviours of the cells and their likely relationships to metastasis

7.1.2.1 Chemotaxis

To begin with, I looked at the chemotactic responses of the cells to the growth factors PDGF and IGF. These growth facters were chosen because the T15s had already been shown to chemotax towards them (Monypenny, 2003). Furthermore, PDGF and IGF may be important in metastasis, since metastatic cells are thought to exploit the homing pathways other cells use to move through the body, and these may include PDGF and IGF. Also, PDGF is present in the smooth muscle cells of blood vessels, and cell migration towards blood vessels is important for the initial intravasation of the metastatic cascade.

The analysis of chemotaxis only involved cells that had migrated further than 70 μm. This restriction removed non-motile cells, improving the reliability of chemotaxis detection since it is only defined for motile cells. I found that the metastatic cells have a significantly stronger chemotactic response to PDGF/IGF than the non-metastatic cells. The persistence of the cells was also checked, and it was found that none of the treatments applied in this thesis had any effect on persistence. A reason for the difference in chemotaxis may be that there is differential expression of the PDGF receptor or IGF receptor amongst the cell populations. For example, the K2 cells would be unable to chemotax if they express fewer receptors for PDGF/IGF. However, the microarray analysis did not indicate any differential expression of the receptors, so this is unlikely to be the cause of the difference in chemotaxis. It is possible that differences in intracellular signalling pathways cause the K2 cells to respond less well to the growth factors, although this needs to be explored further. It has also been proposed that the lack of chemotaxis in the K2 cells is the result of their inability to release an autocrine factor which is involved in chemotaxis (Zicha and Dunn, 1995).

What are the implications of enhanced chemotaxis? A role for chemotaxis in intravasation is supported by a report from (Wyckoff et al., 2000) who observed that metastatic cells can orient themselves towards blood vessels *in vivo*. This relates to chemotaxis since polarisation is the first stage of chemotaxis. There is also evidence that the growth factors employed in this study are specifically involved in some types of metastasis. For example, PDGF receptor autocrine signalling is thought to promote

metastasis in some types of mammary carcinoma (Jechlinger et al., 2006). There is evidence that IGF is involved in invasion of melanoma cells (Girnita et al., 2006), and IGF was also one of the genes in the Amsterdam signature of poor prognosis (van 't Veer et al., 2002). It seems possible in the light of this and other evidence that the poor performance of the K2 cells in the chemotaxis assay contributes to their inability to metastasise.

7.1.2.2 Speed of migration

Differences in speed between the cells were also observed. The K2 cells migrate at approximately 11 μm/hour whereas the metastatic T15 and A297 cells migrate significantly faster, at 17 and 19 μm/hour. Since there is no significant difference in speed between the K2 cells and the A311 cells, which possess the lowest and highest metastatic potential of the model, so there seems not to be an association between speed and malignancy in this model. We had expected that there would be an association, since increased speed is thought to be associated with malignant cells. For example, by performing intra-vital imaging of the tumour cells, Sahai et al. (2005) found that metastatic cells had an enhanced speed compared to non-metastatic cells. It is possible that the cells investigated in this thesis might migrate differently in three dimensions; the cell migration studies described here were performed in an essentially artificial, two-dimensional environment. Future work on this model of metastasis would feature three-dimensional studies *in vitro* and *in vivo*.

It has been suggested that subpopulations of cells from tumours become metastatic, i.e. that not all tumour cells have metastatic potential (Fidler and Hart, 1982). Therefore, we conducted a detailed analysis of the distribution of speeds within the cell populations. This showed that the metastatic cells contain a significantly greater fraction of fast-migrating cells than the non-metastatic cells. It is possible that the increased speed of these subpopulations can contribute to the metastatic potentials observed in the T15, A297 and A311 cells, but this aspect of the study requires further investigation.

7.1.2.3 Actin organisation and dynamics

There are also differences in F-actin organisation between the non-metastatic K2 cells and the metastatic T15, A297 and A311 cells. The K2 cells typically contain stress fibres, while the majority of the metastatic cells in this model do not exhibit stress fibres, but rather have their F-actin arranged into a diffuse network of short filaments or

concentrated into a lamellipodium. Quantification of the images revealed that 68% of K2 cells contain at least one stress fibre compared to only 7.5-10% of the metastatic populations. How do the observed actin arrangements relate to the situation *in vivo*? Motile responses can be made more efficiently from actin that exists as monomers or short oligomers, compared to actin that is largely arranged into stress fibres. This is because stress fibres need to be depolymerised before they can be remodelled into other actin-based structures. The ability to remodel actin has advantages for the metastatic cell. Actin remodelling is known to be an important feature in migration and invasion (Rao and Li, 2004) both of which are essential for metastasis. For example, cortical actin contraction, formation of a lamellipodium, and tail retraction would probably all have to be carried out by an invading cell, and all of these activities are the result of actin remodelling.

How do the observed differences in stress fibres relate to cell adhesion? If there are fewer focal adhesions, the number of stress fibres will be reduced since stress fibres are anchored in focal adhesions. Therefore, it is possible to interpret the loss of stress fibres as a loss of focal adhesions. This would result in an increased ability of the cell to migrate, since focal adhesions do not have to be turned over. This possibility compliments data that has already been obtained on the K2 and T15 cells; namely that K2 cells have thick stress fibres and large focal adhesions, and that T15 cells have thinner actin fibres and smaller focal adhesions (Pokorna et al., 1994). Had there been more time for this project, I would have investigated focal adhesions in detail, since they are likely to be central to the functional effect of 4.1B.

Table 7 summarises the characterisation of the cells and the way in which the behaviours of the metastatic cells might contribute to their metastatic potential.

Table 7: Behaviours of the metastatic cells and the ways in which these might contribute to metastatic potential.

Observation	Relationship to *in vivo* situation
Enhanced chemotaxis to PDGF/IGF.	Polarisation and chemotaxis towards blood vessels are required for intravasation. Hijacking of physiological homing pathways is thought to increase metastasis.
Enhanced migration.	Faster and more efficient completion of the metastatic cascade.
Fewer actin stress fibres.	Actin can be remodelled more quickly and cells are primed for motile responses.
Faster actin dynamics.	Actin can be remodelled more quickly and cells are primed for motile responses.
Fewer focal adhesions.	The cell may be less firmly attached to the substrate, meaning that it can migrate more quickly.

7.2 *Investigation of the gene expression underlying the differences observed between the cells*

The rationale for the microarray experiment was that the differences described above were the result of differential gene expression. The microarray analysis began with the analysis of two cell populations and their responses to growth factors, and was then extended to include a further two cell populations plus primary tumours generated after subcutaneous injection of the cell populations into inbred rats.

Three replicates were performed for each sample, to improve the precision of the estimates of gene expression. The statistical significance of each expression ratio was determined by a t-test. The P-value was corrected using the Benjamini-Hochberg algorithm (Benjamini and Hochberg, 1995) which was necessary to avoid the excess false-positive calls which occur in multiple testing. This helped to produce a final dataset whose reliability was verified in two ways. Firstly, RT-PCR of a sample of candidate genes replicated the expression patterns found by the microarray analysis. Secondly, some of the genes that were identified, for example, *cspg* and *neo1,* have been implicated in metastasis by other groups (Faassen et al., 1993; Kiewe et al., 2006; Lee et al., 2005).

7.2.1.1 Microarray analysis of K2 and A297 cells

The initial microarray experiment, comparing the non-metastatic K2 and 90% metastatic A297 cell populations, and their responses to the growth factors PDGF/IGF, found that 294 genes were differentially expressed between the two cell populations. This is fewer than are typically obtained from similar studies. For example, Gildea et al. (2002) used a human bladder cancer cell line and its isogenic variant to investigate the gene expression of metastasis, and found that 2368 out of the approximately 33000 genes probed were differentially expressed, i.e. 7.2% of genes. In my study, however, only 1.9% of the genes were shown to be differentially expressed. Similarly, my microarray analysis revealed fewer differentially expressed genes than patient-based studies. For example, in the seminal paper by van't Veer et al. describing gene expression associated with poor prognosis, 20% of the transcripts probed by the microarrays were found to be differentially expressed (van 't Veer et al., 2002), almost 10 times the percentage of differentially expressed genes reported here. The reasons that a small number of differentially expressed genes is considered positive are that it is

easier to work on a limited number of genes and that the chances of finding a biologically relevant gene are increased. The small proportion of differentially-expressed genes strongly supports the fitness of the inbred rat sarcoma model for microarray analysis.

7.2.1.2 Effect of PDGF/IGF on sarcoma cell gene expression

A growth factor treatment had been included in the microarray experiment since, in this model, the progression to the invasive phenotype correlated with enhanced chemotaxis in response to a gradient of PDGF/IGF, and I was interested in the transcriptional activity that occurred during this response. However, the microarray unexpectedly revealed very few significant gene expression differences resulting from up to three hours treatment with growth factors. Even when permissive candidate gene criteria were applied, very few genes were seen to be changed in response to growth factors, as expressed in Figure 4.1. There are two possible explanations for this. Firstly, a short, single treatment of PDGF/IGF may not be sufficient to alter gene expression, and may instead exert its effect via molecular switches, such as G-proteins, whose activation state cannot be determined by microarray analysis. Secondly, in the microarray experiment, the cells were exposed to a uniform concentration of growth factors. Cells exposed to the uniform growth factor environment might respond differently compared to cells experiencing a concentration gradient such as that in the Dunn chemotaxis chamber. They may also not respond at all.

An alternative way to explore the effect of PDGF/IGF on gene expression would be to expose cells to a concentration gradient in the Dunn chamber, harvest the cells, amplify the mRNA and perform gene expression analysis. Gene expression analysis on a small number of cells has been carried out successfully by other groups (Wang et al., 2003). However, this is unlikely to be as useful as the exploration of the data already obtained. Firstly, because growth factors probably exert an effect via intracellular signalling leading to the activation of proteins through, for example, phosphorylation events rather than a transcriptional change. Secondly, because other growth factors, for example, EGF, now appear to be more critical *in vivo* and perhaps merit investigation more than PDGF/IGF. For example, the EGF receptor (EGFR) is often expressed in cancer and is associated with a poor prognosis (Klijn et al., 1994).

7.2.1.3 Extension of the experiment

The microarray experiment was extended to include the 40% metastatic T15 cells and 100% metastatic A311 cells, enabling us to look for correlations across a panel of four cell lines from the same model. There were no changes in gene expression in response to the growth factors, as before. We took advantage of this by averaging the mean expression values for the three growth factor treatments. We are certain that it did not distort the results, since our data had shown that there was no differential gene expression following growth factor treatment. Averaging the growth factor treatments improved the quality of the data, because the number of replicates used to obtain the gene expression values was increased. This is borne out by the fact that in the first experiment, one of three genes tested did not have the same expression profile in RT-PCR, whilst in the extended experiment, in which the growth factor-treated samples were pooled, the RT-PCR and microarray data indicated the same expression patterns.

Another improved aspect of the extended microarray experiment is the analysis of primary tumours, which were generated by subcutaneously implanting the cells into the rats. The tumours were subjected to microarray analysis in order to assess whether the gene expression patterns in the cultured cells were due solely to the influence of cell culture conditions. The correlation between the gene expression patterns shows that the gene expression of the cultured cells is an acceptable reflection of the gene expression patterns *in vivo* (see graphs displayed in Figure 4.4 and Appendix 4).

For the primary tumour analysis, a whole section of tumour tissue was sampled. This section would contain blood vessels and stromal cells including any number and type of immune cells. It is possible that these interstitial cells may mask gene expression in the tumour cells themselves. Other groups use techniques such as microneedle collection or laser capture microdissection to avoid this problem (Wang et al., 2003; Yang et al., 2006). Unfortunately such techniques were not within the scope of this project.

The gene expression patterns in the tumours were the same as those in the cultured cells. Therefore it seems that the gene expression of the stromal cells did not appreciably mask the gene expression of metastasis, so I would regard the use of unpurified sections as satisfactory. However, it would still be interesting to see whether certain genes were differentially expressed at the periphery of the tumour compared to the main tumour mass, and whether in fact there are distinct subpopulations of cells

within the tumour, since particular subpopulations of cells may be critical to the behaviour of the tumour (Fidler, 2002).

7.2.1.4 Selection of candidate genes

The rat genome is relatively poorly characterised, meaning that many genes on the array were unannotated. In the final candidate gene list, nine out of the twenty-three genes were unannotated. Therefore, the choice of candidate genes was influenced by the limited availability of reliable sequence data and gene annotation in the rat. It would be interesting to investigate the unannotated candidate genes when more information is available about them.

The annotated genes were associated with a number of GO terms (see Table 6 for the three main biological process GO terms) ranging from embryonic patterning to proteolysis. Some of the genes, for example a*damts1*, which is assigned the GO terms proteolysis, integrin signalling and cell proliferation, have easily imaginable roles in metastasis, which lends support to the microarray data and the approach used here, whilst for other genes, for example, *cask*, whose GO terms are protein complex assembly, protein phosphorylation, and transport, a role in metastasis is not immediately obvious.

Another way to assess the potential relevance of candidate genes is to look at them in the context of their interacting partners or in the context of a gene network (Brazhnik et al., 2002). We coupled the microarray data with a specially developed protein interaction database to identify protein communities associated with metastasis (Jonsson et al., 2006). We found that the majority of communities had already been associated with metastatic behaviour to some degree; for example, MMPs, TGF-β and actinin. This work still has to be validated experimentally, for example, by testing whether the expression or depletion of communities or their regulators has an effect on the behaviour of the cells. Nevertheless the methodology developed may be useful in future for the analysis of microarray data.

7.2.1.5 Comparison of candidate genes with those proposed by other studies

First I will compare my microarray data with that of others, before going on to speculate on the potential relationship of my candidate genes to metastasis. I looked at candidate gene lists obtained from other microarray studies of metastasis to check the universality of the results. Manual inspection of thirteen other microarray-based analyses of

metastasis revealed very little overlap between candidate genes, and this is summarised in Table 8. Of the thirteen studies inspected, only four contained genes overlapping with the genes reported here; in all, five overlapped; *actn*, *epb41l3*, *grem1*, *RAR* and *tpm1*. The small overlap between the datasets is disappointing, but perhaps to be expected given that lack of correlation between data-sets is an emerging feature and challenge of gene expression studies of metastasis (Ein-Dor et al., 2006).

Why is there so little overlap between the studies? Microarray analyses on cohorts of patients have been shown to have a lack of transferability to other data-sets (Michiels et al., 2005). These problems have been attributed principally to the use of too few samples given that the majority of samples come from hosts with heterogeneous genetic backgrounds. Similarly, genetic heterogeneity within animal models might mean that a high proportion of candidate metastasis genes are actually unconnected with metastasis. Additionally, the genetic background of the model, whether heterogeneous or not, will have a certain bearing on the candidate genes resulting from the study. Therefore, the lack of overlap between my and other studies is disappointing but not overly surprising. The implications are that the relevance of the data can only really be ascertained by functional studies. As more and improved microarray datasets become available, the genes frequently occurring in candidate gene lists might be reliably considered to be important players in metastasis.

Table 8: Comparison of microarray data with other studies on the gene expression of metastasis.

Study	Approach	Overlapping genes
Clark et al.	Nude mouse model of metastasis.	α-Actinin up in metastasis.
Gildea et al.	Comparison of human bladder cancer cell lines.	None.
Goswami et al.	Profiling of chemotactic cells isolated from tumours.	Adamts up in metastasis.*
Hao et al.	Comparison of primary breast carcinomas and their lymph node metastases.	None.
Khanna et al.	Mouse model of osteosarcoma	Ankyrin down in metastasis.*
Liu et al.	Nude rat model of metastasis.	epb41l3 down in metastasis.
Minn et al.	Nude mouse model of metastasis.	None.
Pawitan et al.	Analysis of 159 primary breast carcinomas.	None.
Ramaswamy et al.	Comparison of primary/secondary human adenocarcinomas.	Gremlin and RAR down in metastasis.
van 't Veer et al.	Analysis of 117 primary breast carcinomas.	None.
Wang et al. (2002)	Rat model of metastasis.	None.
Wang et al. (2004)	Analysis of chemotactic breast cancer cells.	Tropomyosin and α-Actinin up in metastasis.
Yanagawa et al.	Comparison of primary colorectal tumours with liver metastases.	None.

** indicates that the expression pattern reported is the opposite of that found in this thesis*

7.2.1.6 Potential roles of the candidate genes in metastasis

The candidate genes *bk*, *cask*, *epb41l3* and *ril* will be discussed in the next section since they were subjected to functional studies. Here I speculate upon the relevance of some of the other candidate genes.

Grem1, coding for Gremlin (also called DRM), was found to be downregulated with metastasis both by this study and by Ramaswamy et al. (2003). *Grem1* was found to be repressed in transformed rat embryonic fibroblasts and its exogenous expression therein causes apoptosis (Topol et al., 1997). Therefore, its expression in the K2 cells may prevent their successful growth in a new site and account for their inability to cause metastases. Gremlin also functions as a negative regulator of monocyte chemotaxis (Chen et al., 2004). The expression of a negative regulator of chemotaxis in the non-metastatic cells may explain their poor chemotactic responses to PDGF/IGF.

Adamts1, encoding the metalloprotease ADAMTS-1, is downregulated with metastasis in this model. Our data agree with that of Kuno et al. (2004) who report that overexpression of ADAMTS-1 in CHO cells causes inhibition of tumour growth and metastasis. It is counter-intuitive that a metalloprotease would be downregulated with metastasis, since degradation of ECM is known to be important in cancer, however, tumour cell migration may also occur in a non-proteolytic manner (Friedl and Wolf, 2003) meaning that proteases would not necessarily have to be expressed. *RNase 4* is also downregulated with metastasis. Although there is currently no clear role for *RNase 4* in cancer, unusual levels of RNase have been detected in cancer patients (Peracaula et al., 2000) and *RNase 4*, has been implicated in prostate cancer (Silverman, 2003). *Neo1* (Neogenin) is also downregulated in metastasis in this model. Its expression has been found to be inversely correlated with mammary carcinogenicity (Lee et al., 2005) which would agree with my data.

Cspg4, which encodes the proteoglycan NG2, is overexpressed approximately 65-fold in metastatic cells, and 14-fold in metastatic tumours. This agrees with data from Burg et al. (1998) who reported that expression of NG2 enhanced the metastatic properties of melanoma cells. NG2 is a proteoglycan involved in cell adhesion and its increased expression is likely to confer an advantage on cells during the metastatic cascade, by, for example, enhancing their ability to adhere to the vessel wall during extravasation. Ankyrin 3 is upregulated approximately 19-fold in the metastatic cells and tumours.

Ankyrins link the cell membrane to the spectrin-based cytoskeleton and are thought to be involved in migration. Bourguignon et al. (2000) report that the interaction of Ankyrin with Tiam1 promotes migration in mouse breast tumour cells. The function of Tiam1 is likely to be affected by an increased availability of Ankyrin. Therefore, an increase in Ankyrin expression, as occurred in the metastatic cells of this model, could lead to enhanced invasion, which is important for metastasis. It is also possible that the increase in Ankyrin expression accounts in part for the enhanced speed of the T15 and A297 cells of this model.

7.3 Functional studies of selected candidate genes

Here I will discuss the functional studies carried out on candidate genes *actn, bk, cask* and *ril. Epb41l3* will be discussed separately in Section 7.4, since many implications arose from the experiments on *epb41l3*.

7.3.1 α-Actinin

My results, showing that α-Actinin is upregulated in metastatic A297 cells, are similar to those obtained in a microarray analysis of pulmonary metastases against parental cell line grown as subcutaneous tumours, where α-Actinin was enhanced between 3.3-7.3 times in metastases compared to their corresponding primary tumours (Clark et al., 2000). The cellular localisation of α-Actinin was investigated by immunocytochemistry and colocalisation was seen with actin in both cell types, as expected. There was slightly less α-Actinin staining in K2 cells which supported the expression pattern seen in the microarray data. Neither metastatic cells overexpressing α-Actinin nor non-metastatic cells with artificially depeleted α-Actinin displayed any difference in motility. This is understandable in view of the fact that α-Actinin is not on the final gene list; the expression ratio and statistical significance were not particularly high in comparison to some candidate genes, and in view of the extended dataset, may not represent significant expression changes but rather random or redundant fluctuations in gene expression. In view of this new information, α-Actinin was not explored further, although it served as a useful pilot study for how the functional studies would take place.

7.3.2 BK

The gene *bk* was upregulated in metastatic cells in this model. Its protein product BK belongs to a family of C2 domain-containing proteins, other members of which are synaptotagmin and rabphilin 3A (Fukuda and Mikoshiba, 2001; Kwon et al., 1996). Its name derives from the fact that it was initially thought to be expressed selectively in the brain and kidney, although it is now known to be expressed elsewhere (Fukuda and Mikoshiba, 2001). It has been shown to target to the membrane via an N-terminal cysteine cluster, and Fukuda and Mikoshiba (2001) speculate that this motif may allow it to localise to the golgi apparatus. The function of BK is unknown.

There are no antibodies against BK available. This meant that firstly, I was unable to confirm the protein expression by immunoblot. Although I have ascertained the expression of the Bk gene by both microarray and RT-PCR, it is possible, though somewhat unlikely, that the transcriptional activity does not mirror the translational activity in the cells, as the mRNA may be preferentially translated in different tissues or at different stages. Secondly, I had to rely upon cDNA constructs of BK to investigate its localisation in the cells.

I cloned and expressed GFP-BK in the sarcoma cells. The BK protein was C-terminally tagged since I did not wish to place the GFP tag at the N-terminus and risk disrupting the membrane targeting which is believed to occur via the N-terminal cysteine cluster. When GFP-BK was expressed in Hela cells, it had a wide cellular distribution with a slight accumulation in the nucleus, and was not significantly enriched at the membrane, although it is reported to be tightly bound to the membrane fraction (Fukuda and Mikoshiba, 2001). This raised the possibility that the BK protein may not have been correctly expressed or folded by the cells, or that the GFP tag interfered with the membrane localisation and activity of BK. I therefore removed the GFP tag from BK for the subsequent motility studies.

I found no significant difference between the motility of cells injected with GFP and cells injected with the mixture of BK/GFP. This indicated that BK has no effect on motility in the sarcoma cells. However, without a functional antibody, it was not possible to assess whether BK was being expressed and localised correctly in the cells. Similarly, without an antibody to check RNAi efficiency, I was unable to perform the

complimentary motility experiments in the metastatic A297 cells, which are believed to overexpress BK.

In summary, technical limitations opened these functional studies to uncertainty. For example, I could not check the possibility of a discrepancy between the mRNA and protein expression, nor could I check the cellular localisation of the endogenous or ectopic protein. Therefore I decided to abandon the experiments on BK. If time had allowed, I would have generated an antibody against BK in order to resolve the issues mentioned above, before beginning overexpression and knock-down studies. An alternative way to check the correct targeting of the BK construct would be to fuse it to a FLAG or similarly small tag. FLAG tag technology consists of tagging a protein with an eight amino acid motif, which is very likely too small to interfere with protein function. The protein is then expressed by the cells and the cells are fixed and probed with an anti-FLAG antibody, to check the expression and localisation.

7.3.3 CASK

The gene *cask* and its protein product are upregulated in K2 cells compared to the three metastatic cell populations. CASK was identified as a neurexin-interacting protein (Hata et al., 1996) and is composed of a CaM Kinase domain, a DHR domain, PDZ domain, SH3 domain and a guanylate kinase domain. The structure suggests that CASK is an adaptor protein, presumably localising to the membrane via its SH3 domain and possibly mediating calcium signalling events via its CaM kinase domain. CASK has also been shown to bind to protein 4.1, which suggests a role in linking the ECM and the cytoskeleton (Cohen et al., 1998).

The protein expression was confirmed by immunoblotting, and the cellular localisation was investigated in fixed cells using the same antibody. The staining pattern of the CASK antibody was diffuse throughout the cytoplasm and there was no increase in CASK staining in the K2 cells. In fact, it was possible to see a slight increase in CASK staining in the metastatic cells which is the opposite of what would be expected given the expression patterns revealed by the microarray, RT-PCR and immunoblot. Since the antibody had worked well in immunoblotting (i.e. without recognising non-specific proteins) it seems reasonable to conclude that the antibody simply does not work well for immunocytochemistry and can only be relied upon for immunoblotting. This is likely due to the manner in which the antibody was generated. For example, if a peptide

was used to make an antibody, it is possible that the peptide site was exposed by the procedures for immunoblotting but not by the procedures for immunocytochemistry. The amount of CASK probably appeared to be greater in the A297 with immunocytochemistry because of cross-hybridisation.

I went on to investigate the effect of overexpression of GFP-CASK on the sarcoma cells. The GFP was tagged to the N-terminus of the protein so as to not interfere with the putative membrane binding. GFP-CASK could be detected on an immunoblot by the anti-CASK antibody, which indicated that CASK was expressed correctly by the cells. Expression of GFP-CASK in A297 cells was seen throughout the cytoplasm and nucleus and there was no colocalisation with F-actin and no enrichment at the membrane. It is possible that this is the correct localisation of CASK, and that it only binds the membrane when the correct calcium signals are received and transduced by the CaM kinase domain. If more time had been available, I would have improved the investigation of CASK by observing the behaviour of the cells in response to activation of the CaM kinase cascade (Soderling, 1999). This might reveal an effect on cell behaviour which is not evident in basal conditions.

Again, it is difficult to be sure that the GFP tag (which was beside the CaM kinase domain) did not interfere with the function of CASK, and if I had more time to investigate this aspect, I would use a FLAG-tagging strategy to try and establish its cellular localisation. For the motility studies, untagged CASK was used. I found no difference in speed between A297 cells injected with GFP/CASK and those injected with GFP alone. I then performed the complimentary experiment in K2 cells by reducing CASK expression by RNAi. Again, there was no difference in speed between cells expressing different levels of CASK, nor was there any difference in cell morphology or actin cytoskeleton in CASK RNAi A297 cells. Since these complementary approaches had no effect on the cells it is clear that CASK alone cannot be responsible for the differences in cell motility and cytoskeleton.

Although I have not been able to show experimentally that CASK has a role in metastasis or any associated behaviours, work by others suggests that CASK may be involved in metastasis. For example, Al-Lamki et al. (2005) investigated the genes associated with relapse in paediatric acute lymphoblastic leukaemia and found that overexpression of CASK was common in non-relapsed patients. It has also been

reported that CASK was upregulated in 87% of human oesophageal carcinomas tested by Northern analysis (Wang et al., 2002a). This would agree with my data that shows that CASK is overexpressed in non-metastatic sarcoma cells.

7.3.4 RIL

Reversion-induced LIM protein (RIL) is overexpressed with metastasis in this model. RIL contains a PDZ and an LIM domain, and was found to be upregulated in Jun-transformed cells (Fu et al., 2000). RIL may be a regulator of actin stress fibre turnover, since the expression of RIL causes the rapid formation of new stress fibres and frequent collapse of thick stress fibres (Vallenius et al., 2004). As RIL is upregulated in metastatic cells which do not have thick stress fibres, one might speculate that the presence of RIL contributes to the turnover of filaments and the phenotype of reduced actin stress fibres. Similarly, the upregulation of RIL in the metastatic cells would account for the faster actin dynamics in T15 cells.

I used a GFP-fusion protein of RIL to investigate its cellular localisation. The colocalisation of RIL with actin stress fibres demonstrated that the protein was expressed correctly and had actin-binding function. The lack of effect on actin stress fibres in K2 cells, which typically do not express RIL, was surprising, as was the absence of a change in speed when the cells were injected with RIL. This indicates that RIL is unable to affect stress fibres on its own. The function of RIL in stress fibre turnover does not seem to affect these cells, although this aspect would be better investigated in live cells, for example by FRAP (fluorescence recovery after photobleaching).

Since RIL contains a PDZ and LIM domain it may act in synergy with α-Actinin, which is a PDZ containing protein. If the two proteins were co-expressed, they may have some effect on the cells and this is something that regretfully I was unable to explore. Vallenius et al. report that RIL can enhance the association of α-Actinin with F-actin. If time had allowed, I would have investigated whether the coordinate deregulation of these two proteins is able to elicit a change in the cells.

7.4 Is there a role for 4.1B in metastasis?

The candidate gene *epb41l3* is significantly down-regulated in metastasising sarcoma cells. The similarity of expression patterns in the primary tumours and the cultured cells, shown in Figure 4.4, excluded the possibility that the expression pattern of *epb41l3* was solely a result of cell culture conditions. Immunoblotting showed that the protein product 4.1B was also downregulated in the metastatic cells.

7.4.1 4.1B as a tumour suppressor

The tumour suppressor activity of 4.1B has been demonstrated in lung, meningioma and breast cancer cell lines (Charboneau et al., 2002; Gutmann et al., 2001; Tran et al., 1999). 4.1B is lost in about 60% of lung adenocarcinomas (Tran et al., 1999), meningiomas (Gutmann et al., 2001) and renal clear cell carcinomas (Kikuchi et al., 2006). However, Yi et al. (2005) recently developed a 4.1B deficient mouse and found it to be indistinguishable from wild-type mice in terms of development and tissue morphology. Furthermore, the $4.1B^{-/-}$ mice had no change in cellular proliferation and no predisposition to cancer, under the conditions tested. This raised the possibility that 4.1B loss occurs in association with, rather than as a cause of cancer (Yi et al., 2005). However, the lack of apparent difference in cancer predisposition in $4.1B^{-/-}$ mice does not rule out a causal role in cancer, as other factors, such as genetic background and functional compensation, must also be considered. The effect of 4.1B loss would be dependent upon the genetic background of the model system or of the patient, just as the effect of any other alteration is influenced by the genetic background in which it occurs. Furthermore, the absence of 4.1B in this mouse may be compensated for by other proteins, such as other members of the 4.1 superfamily or splice variants of 4.1B. Finally, 4.1B may be more important in the progression of cancer rather than its initiation.

7.4.2 A new epb41l3 splice variant

I found that the *epb41l3* transcript in the sarcoma cells is a variant of the 4.1B minor isoform type II, which represents a third splice variant of *epb41l3*. This variant was submitted to GenBank and assigned the Accession Number DQ462202. By checking the predicted protein sequence of the new isoform, I confirmed that the protein domains that characterise 4.1B were conserved. The discovery of a third splice variant underscores the importance of alternative splicing in the effects of 4.1B and the 4.1 superfamily. Different splice variants of *epb41l3* are thought to contribute to its tissue specificity (Gascard et al., 2004). As yet the function of the new splice variant is not known.

7.4.3 What is the cellular localisation of 4.1B?

DAL-1 had been localised to cell-cell contacts (Tran et al., 1999) and to the plasma membrane (Charboneau et al., 2002). Although the 4.1B antibody worked satisfactorily for immunoblotting, it did not work in immunocytochemistry, in a range of different fixation and staining conditions. Therefore, the only way to localise the protein was with the use of genetic fusions. Constructs encoding a 4.1B protein tagged at the N- or C-terminus with GFP were generated and expressed. Co-staining with rhodamine phalloidin was performed, since 4.1B contains an SAB domain, known to bind actin, and I wished to check whether GFP-4.1B would colocalise. Expression of GFP-4.1B was diffuse in the cytoplasm with no expression in the nucleus. Some colocalisation with actin was observed, as indicated by arrow heads in Figure 6.6. There were areas of higher expression at the edge of the cell, as indicated by the arrowheads in Figure 6.8B and the high magnification image in Figure 6.8C. This agrees with an observation by Tran et al. who reported minor punctuate staining at the cell edges, and Charboneau et al., who observed an enrichment of 4.1B at the cell membrane. To avoid the potential problem of the GFP interfering with the protein function, an untagged version of 4.1B was used in the subsequent functional studies. Incidentally, the problems associated with tagging FERM domain-containing proteins have been commented on before (Chambers and Bretscher, 2005). Had more time been available, I would have employed a FLAG-tagging strategy, or preferably generated another antibody to 4.1B, to check more rigorously the localisation of the protein.

7.4.4 Does 4.1B regulate cell motility and the actin cytoskeleton?

The effect of depleting expression of 4.1B in K2 cells, in which it is endogenously overexpressed, was assessed. K2 cells microinjected with siRNA against *epb41l3* displayed an altered F-actin cytoskeleton, with fewer stress fibres and less total F-actin than cells microinjected with control siRNA. Similar results were obtained in HeLa cells, illustrating that 4.1B loss can cause an effect in more than one cell type. The stress fibre phenotype could be restored by expressing siRNA-resistant *epb41l3* in the cells. This was achieved by the use of an untagged rat *epb41l3*; the siRNA used for the HeLa cells was targeted against sequences specific to human *epb41l3* and not conserved in rat *epb41l3*. The recovery of stress fibres is important since it indicates that all three siRNA were correctly targetting the *epb41l3* sequence and that the loss of stress fibres is a specific result of 4.1B loss. The specificity of the siRNA was further backed up by experiments in which siRNA against *epb41l3* was shown to elicit a phenotypic change in the K2 cells and not the A297 cells, the reason being that the A297 do not express *epb41l3*, the target transcript of the siRNA.

Metastatic cells in this model contain fewer stress fibres and less 4.1B than the non-metastatic cells. Similarly, cells made to express less 4.1B exhibit fewer stress fibres. Taken together, these observations suggest that the loss of 4.1B in the metastatic cells is partly responsible for their lack of stress fibres. Interestingly, a reduction in stress fibres accompanying the metastatic phenotype has been observed by others (Khanna et al., 2001).

The relationship between 4.1B and the actin cytoskeleton was further explored by expressing 4.1B in the metastatic cells. The ectopic expression of 4.1B in metastatic T15 cells, which do not endogenously express 4.1B, had no apparent effect on the arrangement of F-actin. There are a number of reasons that can account for the non-appearance of stress fibres in T15 cells overexpressing 4.1B. As seen from the microarray, a wide range of genes are differentially expressed between the non-metastatic and metastatic cells. These additional proteins may be required by the metastatic cells in order to form a significant number of stress fibres. Since 7-10% of the metastatic cells do contain stress fibres there must be other stress fibre-forming factors at work besides 4.1B. One of the differentially-expressed proteins may be a key binding partner of 4.1B, meaning that overexpressed 4.1B would have no opportunity to affect the cytoskeleton. Alternatively, 4.1B may have no effect on the actin

cytoskeleton because, for example, its signalling partners are being used by another protein. The third and perhaps most likely possibility is that the method that I used to detect the stress fibres is not sensitive enough to detect 4.1B-induced differences in the T15 cells; I used a binary classification where cells were scored as having stress fibres or no stress fibres. Automatic quantification of the cells on the basis of their rhodamine phalloidin signal might reveal differences between cells expressing 4.1B and control cells, although the natural heterogeneity of the sarcoma cells renders such a strategy very difficult.

As 4.1B RNAi disrupted the actin cytoskeleton of K2 cells, its effect on cell migration was also tested. K2 cells were microinjected with control or *epb41l3* siRNA and their migration recorded 48 hours later. Reduction of 4.1B expression by siRNA treatment caused cell speed to double. Although 4.1B expression did not induce the formation of stress fibres in the metastatic T15 cells, it reduced their speed by half. The fact that overexpression of 4.1B as well as its depletion can lead to changes in speed in different cell types, strongly supports the possibility that 4.1B is important for cell migration in this model.

7.4.5 How does 4.1B affect stress fibres and speed?

The 4.1B RNAi-induced loss of stress fibres and increase in cell speed is striking because these two phenotypes also occur in the metastatic cells, which do not endogenously express 4.1B. Furthermore, actin rearrangements accompanying changes in speed have been observed by others. For example, Clark et al. (2000) reported that metastatic cells migrate more and have more filopodial protrusions than cells from the primary tumours. The link between loss of stress fibres and increase in speed may be that cells in which actin is not polymerised into stress fibres are primed for migration since their largely monomeric actin can be more easily remodelled into the structures required for motility, as mentioned in Section 7.1.

Since stress fibres are localised at focal adhesions (Parsons et al., 2000) the presence of 4.1B in focal adhesions might explain its affect on the actin cytoskeleton. Perhaps a clustering of 4.1B within the focal complexes provides an anchor for actin filaments and helps them to remain organised in stress fibres. There is some evidence to suggest that 4.1B is involved in focal adhesions. Firstly, 4.1B has been shown to colocalise with the focal adhesion protein Paxillin in cultured astrocytes (McCarty et al., 2005). Secondly, 4.1B is thought to affect cell adhesion. It has been observed that DAL-1/4.1B-expressing breast carcinoma cells were better able to adhere to Fibronectin or Matrigel™ than DAL-1/4.1B non-expressing cells (Charboneau et al., 2002), and Gutmann et al. (2001) observed spreading defects in RT4 schwannoma cells overexpressing DAL-1. Thirdly, 4.1B can interact with proteins that are known to be involved in adhesion, such as integrin αvβ8 (McCarty et al., 2005) and TSLC-1 (Yageta et al., 2002). Together these reports would support the idea that loss of 4.1B might disrupt the focal adhesions. If 4.1B is involved in maintaining focal adhesions, it may be the effect on these, rather than a direct effect on stress fibres, that causes the phenotype reported in this study.

The regulation of focal adhesions can also affect cell speed. For example, FAK-deficient mice, which have increased focal adhesions, also display reduced motility (Ilic et al., 1995). Therefore, disruption of focal adhesions by perturbation of 4.1B levels could account for the increase and decrease in cell speed that occurs upon 4.1B depletion and overexpression respectively. For example, overexpressed 4.1B

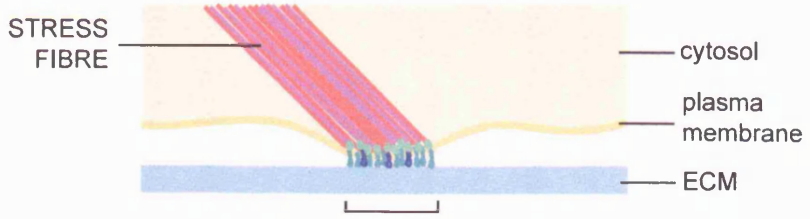
arranged into focal complexes might hinder the de-adhesion that is an essential part of cell migration. Similarly, the reduced amount of 4.1B in focal adhesions might cause the adhesions to be slightly less stable and increase the speed with which they can be turned over by the cell and consequently increase the speed of cell migration. These ideas are summarised in Figure 7.1.

Of the three main domains on protein 4.1B, it is most likely that the observations reported here are mediated by the FERM domain, which has been shown to link the cytoskeleton to the ECM (Chishti et al., 1998). Therefore, reduced availability of the FERM domain of 4.1B may mean that the cell/ECM linkage is lost. This would mean that cells can move in a faster and uncontrolled way. In comparison, cells that overexpress the 4.1B protein move more slowly, since there are more interactions between the cell and the substrate, meaning that adhesion is too high and cell migration is impeded.

Had more time been available, I would have tested the effects of different truncation mutants of 4.1B in order to postitively identify the domain interactions responsible for the changes in cell motility. Given more time I would also characterise the focal adhesions in cells with altered levels of 4.1B. It has however already been shown that the K2 cells (which overexpress 4.1B) have larger focal adhesions than the 4.1B non-expressing T15 cells (Pokorna et al., 1994) and so a correlation if not a direct link between 4.1B and focal adhesions has been shown.

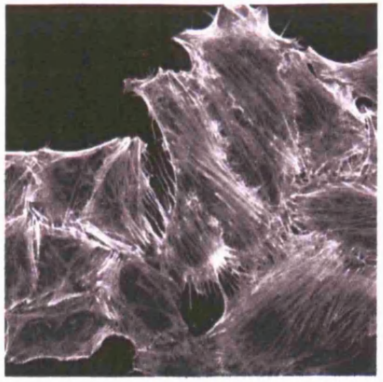
Figure 7.1: What are the consequences for the cell if 4.1B is involved in focal adhesions?

A focal adhesion containing various proteins including 4.1B.

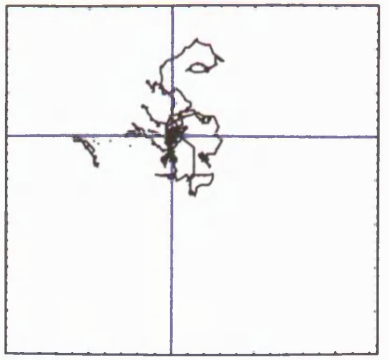


FOCAL ADHESION formed from a variety of proteins, including 4.1B.

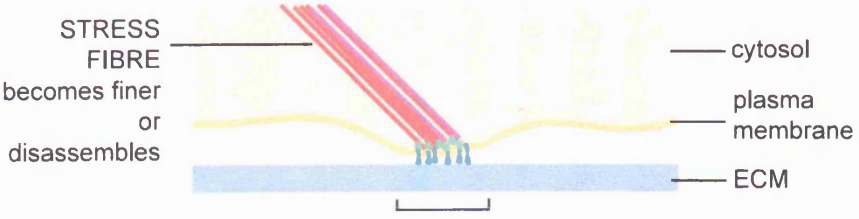
Focal adhesions are strong and stress fibres form.



Focal adhesion turnover is slow and speed is moderate.

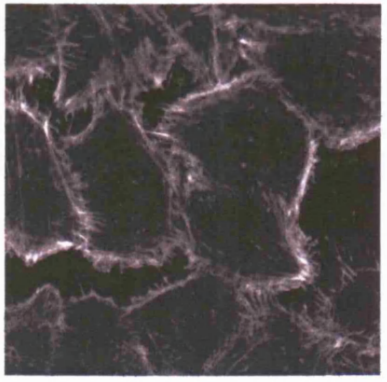


A focal adhesion without 4.1B.

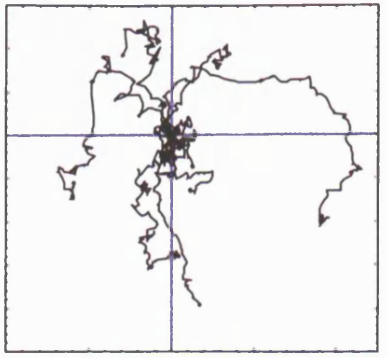


FOCAL ADHESION becomes weaker.

Focal adhesions are weaker and stress fibres are rarer.



Focal adhesion turnover is rapid and speed increases.



The possibility that 4.1B is involved in focal adhesions can explain some the data presented in this thesis.
Upper panel: *4.1B contributes to the focal contacts into which stress fibres are embedded. Stress fibres can be seen in rhodamine phalloidin-stained cells. The speed of the 4.1B-containing sarcoma cells is approximately 11 μm/hour.*
Lower panel: *The absence of 4.1B from the cells, either in metaststic cells or in non-metastatic cells treated wtih RNAi, may result in a destabilisation of the focal contacts which in turn means that fewer stress fibres are present. The weakness of the focal contacts means that they are rapidly turned over by the cell and migration increases.*

7.4.6 What are the other mechanisms by which 4.1B might affect cell motility and the cytoskeleton?

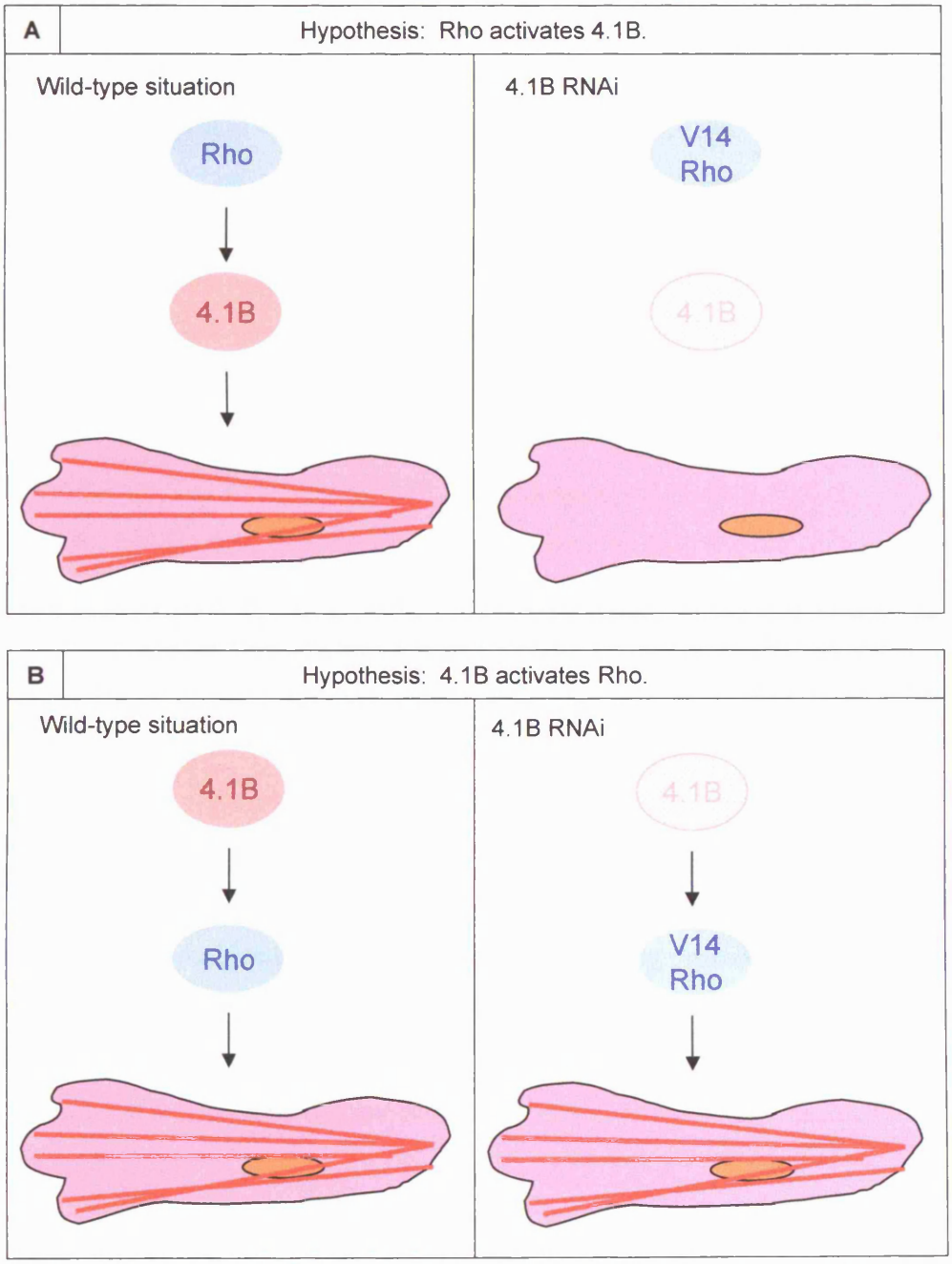
4.1B could affect cell motility and the cytoskeleton in other ways besides contributing to focal adhesions. 4.1B has recently been shown to modulate the methylation activity of other proteins and this ability may affect cell motility and the cytoskeleton. 4.1B interacts with N-methyltransferase 3 (PRMT3) (Singh et al., 2004) and PRMT5 (Jiang et al., 2005), although 4.1B does not become methylated by PRMT3 or PRMT5 *in vitro*, but rather modifies their ability to methylate other protein substrates in the cell. If 4.1B is able to modulate the activity of a PRMT, it may be able to affect the cell in many ways, since methylation of proteins can influence signal transduction and transcriptional activation (McBride and Silver, 2001). The interaction of 4.1B with PRMT3 and PRMT5 may be a mechanism through which 4.1B regulates tumour suppression. Indeed, in MCF-1 breast cancer cells, artificial protein methylation and 4.1B were shown to enhance apoptosis (Jiang and Newsham, 2006).

The mechanism by which 4.1B may stabilise or contribute to the formation of stress fibres is most likely via the GTPase Rho, as Rho controls both stress fibre formation and focal adhesion formation (Ridley and Hall, 1992). Moreover, other ERM proteins have been shown to be required for Rho function (Mackay et al., 1997) so it would seem reasonable to suppose that 4.1B might also affect Rho function. I therefore performed an experiment with constitutively active (V14) Rho in an attempt to determine the signaling involved. If 4.1B were an effector of Rho, V14 Rho would not be expected to form stress fibres in a 4.1B RNAi background (see Figure 7.2A). V14 Rho was found to recover stress fibres in 4.1B RNAi K2 cells. The observation that V14 Rho can make stress fibres independently of 4.1B could be explained by Rho being downstream of 4.1B (see Figure 7.2B). This would mean that the depletion of stress fibres in 4.1B RNAi cells occurs because Rho does not become activated by 4.1B.

This hypothesis conflicts with a recent report (Gerber et al., 2006) that 4.1B has no effect on the activity of Rho. An alternative explanation for the ability of V14 Rho to form stress fibres in a 4.1B RNAi background is that 4.1B and Rho operate through different signaling pathways. Alternatively, the depletion of 4.1B by siRNA treatment may have been insufficient, meaning that 4.1B was still available to be activated by V14 Rho. This aspect of 4.1B action needs to be explored in further detail by, for example,

the performance of Rho pull-down experiments and by testing whether 4.1B is able to form stress fibres in cells expressing the dominant negative N17 Rho.

Figure 7.2: Schematic illustrating the possible interplay between 4.1B and Rho.



Experiments were performed in a 4.1B RNAi background to explore the possible interplay between 4.1B and the small GTPase Rho. In hypothesis A, activation of 4.1B results in the formation of stress fibres. In a 4.1B RNAi background therefore, constitutively active V14 Rho will not form stress fibres. Hypothesis B is that 4.1B activates Rho. If this is the case, V14 Rho will be able to form stress fibres in the absence of 4.1B.

7.4.7 Other evidence for a role for 4.1B in metastasis

A range of studies suggest that 4.1B is a tumour suppressor (Charboneau et al., 2002; Gutmann et al., 2000; Tran et al., 1999). My data suggest that 4.1B may also function as a metastasis suppressor as its loss supports the enhanced cell motility which is required for metastasis (Ahmad and Hart, 1997) and a concomitant reorganisation of the actin cytoskeleton. Had more time been available, stable cell lines with depleted or overexpressed 4.1B would have been generated to test whether the metastatic potentials of the K2 and T15/A297/A311 cells could be reversed. As metastasis assays have not yet been performed, I cannot say with certainty that 4.1B can act as a metastasis suppressor *in vivo*. There is however some evidence to suggest that 4.1B loss is associated specifically with a tendency to metastasise. For example, loss of heterozygosity of the region of chromosome 18 where 4.1B is localized has been observed in more than 55% of invasive DCIS (ductal carcinoma in situ) tumours (Kittiniyom et al., 2001). Also, in a microarray analysis of cancer cell lines, Liu et al. (2004) found that 4.1B is downregulated in metastatic variants of the H460SM lung cancer cell line. Loss of 4.1B via epigenetic inactivation is also associated with metastasis in human cancers. The epigenetic inactivation of 4.1B by promoter methylation was found to be an indicator of poor prognosis in patients with lung adenocarcinoma, as it is associated with a shorter survival (Kikuchi et al., 2005). Preferential methylation of 4.1B in lung adenocarcinomas with invasion to the surrounding tissues, lymph nodes (often the first step in metastatic disease) and pleura was also observed. Similarly, epigenetic inactivation of 4.1B by promoter methylation was also found to correlate with a shorter recurrence-free survival in patients with renal clear cell carcinoma (Kikuchi et al., 2006). These observations support the data I have shown to suggest that loss of 4.1B can enhance cell behaviours which are associated with metastatic potential.

7.4.8 Summary

To summarise this section, 4.1B was found to be downregulated in metatastic cells. Fucntional assays of 4.1B by both siRNA treatment and cDNA expression strongly suggest that 4.1B has a role in both cell motility and the cytoskeleton. The mechanism through which 4.1B exerts its effects remains unknown but a role in metastasis seems likely, as loss or epigenetic inactivation of 4.1B has been associated with invasive cancer and a shorter survival time in cancer patients.

7.5 Conclusion

In this thesis I have described the use of gene expression microarray technology to search for genes that relate to the development of metastasis. The model of metastasis used for the microarray was chosen because of its uniform genetic background and correlation between cell behaviour *in vitro* with metastatic potential *in vivo.* The success of the microarray experiment is demonstrated by the fact that the data are reproducible in other assays, and that some of the candidate genes have already been identified in metastasis. Furthermore, one of the candidate genes, *epb41l3,* has a functional effect in the cells.

Epb41l3, encoding the protein 4.1B, is downregulated in metastatic cells and its reintroduction reduced the speed of the cells by half. Conversely, the depletion of 4.1B from non-metastatic cells resulted in an increase in cell speed and concomitant loss of stress fibres. This is important because both cell migration and the regulation of the actin cytoskeleton are known to be involved in metastasis.

I conclude that the loss of 4.1B in the metastatic cells causes a remodelling of the actin cytoskeleton together with an increase in speed that may promote the ability of the cell to metastasise. Others have associated 4.1B with cancer (Charboneau et al., 2002; Gutmann et al., 2000; Tran et al., 1999) and specifically with human cancers that are invasive or likely to invade (Kikuchi et al., 2006; Kikuchi et al., 2005; Kittiniyom et al., 2001). The clinical implications of this are that 4.1B could be a useful prognostic marker for cancers since its loss may be indicative of metastasis.

7.6 Reflections on the approaches used in this thesis

Since completing the research for this thesis I have reflected on the approaches used. The central approach was the use of a gene expression microarray to shed light on some of the gene expression controlling metastasis. A possible criticism of this work could be that, as 4.1B is thought to be a tumour suppressor, an investigation into its role in metastasis might have been initiated without the use of microarray technology, which is costly and was at the time rather difficult. Whilst this is true, I believe that the microarray analysis that was conducted was a sensible, rigourous approach, and that it cannot be undermined by the fact that the main gene investigated, 4.1B, was already known to be involved in cancer. Secondly, the use of the microarray can be justified by the fact that my data validates other data, which is a valuable contribution in itself. Thirdly, the microarray data has not yet been fully exploited, and it is likely that investigation of other candidate genes may reveal entirely novel, useful information.

With regard to the model that I chose, it has now come to light (Ein-Dor et al., 2006) that many thousands of samples are needed to generate a robust gene expression signature for patients. This is due to the fact that the variation between patients interferes with the collection of reliable data. This information supports my reasons for using the model of metastasis that I did; a model in which the subjects are genetically related, thus supplying little heterogeneity between samples.

The ideal project would have been to collect tissue samples from hundreds of patients and conduct a tissue microarray. However, at the time that my project began, tissue microarrays were not yet particularly reliable. Similarly, the number of samples needed to produce reliable data was simply not known. In terms of the scope of my project and the constraints of time, budget and labour, I think that my approach is justified and provided useful information.

Were I to continue with my research, I would try to exploit my microarray data fully, and compare it to emerging patient data in order to assess the relevance of the genes, before moving on to careful and thorough functional studies.

References

Abercrombie, M. and Heaysman, J. E. (1966). The directional movement of fibroblasts emigrating from cultured explants. *Ann Med Exp Biol Fenn* **44**, 161-5.

Ahmad, A. and Hart, I. R. (1997). Mechanisms of metastasis. *Crit Rev Oncol Hematol* **26**, 163-73.

Al-Lamki, Z., Wali, Y. A., Wasifuddin, S. M., Zachariah, M., Al-Mjeni, R., Li, C., Muralitharan, S., Al-Kharusi, K., Gunaratne, P., Peterson, L. et al. (2005). Identification of prognosis markers in pediatric high-risk acute lymphoblastic leukemia. *Pediatr Hematol Oncol* **22**, 629-43.

Alberts, A. B., D; Lewis, J; Raff, M; Roberts, K; Watson, J. (2002). Molecular Biology of the Cell: Garland.

Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., Davis, A. P., Dolinski, K., Dwight, S. S., Eppig, J. T. et al. (2000). Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet* **25**, 25-9.

Barone, A. D., Beecher, J. E., Bury, P. A., Chen, C., Doede, T., Fidanza, J. A. and McGall, G. H. (2001). Photolithographic synthesis of high-density oligonucleotide probe arrays. *Nucleosides Nucleotides Nucleic Acids* **20**, 525-31.

Bateman, A., Coin, L., Durbin, R., Finn, R. D., Hollich, V., Griffiths-Jones, S., Khanna, A., Marshall, M., Moxon, S., Sonnhammer, E. L. et al. (2004). The Pfam protein families database. *Nucleic Acids Res* **32**, D138-41.

Benjamini, Y. and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Royal Stat Soc* **57**, 289-300.

Black, W. C. and Welch, H. G. (1993). Advances in diagnostic imaging and overestimations of disease prevalence and the benefits of therapy. *N Engl J Med* **328**, 1237-43.

Bogenrieder, T. and Herlyn, M. (2003). Axis of evil: molecular mechanisms of cancer metastasis. *Oncogene* **22**, 6524-36.

Bourguignon, L. Y., Zhu, H., Shao, L. and Chen, Y. W. (2000). Ankyrin-Tiam1 interaction promotes Rac1 signaling and metastatic breast tumor cell invasion and migration. *J Cell Biol* **150**, 177-91.

Bray, D. (2001). Cell Movements from Molecules to Motility.

Brazhnik, P., de la Fuente, A. and Mendes, P. (2002). Gene networks: how to put the function in genomics. *Trends Biotechnol* **20**, 467-72.

Bretscher, A., Edwards, K. and Fehon, R. G. (2002). ERM proteins and merlin: integrators at the cell cortex. *Nat Rev Mol Cell Biol* **3**, 586-99.

Buccione, R., Orth, J. D. and McNiven, M. A. (2004). Foot and mouth: podosomes, invadopodia and circular dorsal ruffles. *Nat Rev Mol Cell Biol* **5**, 647-57.

Burg, M. A., Grako, K. A. and Stallcup, W. B. (1998). Expression of the NG2 proteoglycan enhances the growth and metastatic properties of melanoma cells. *J Cell Physiol* **177**, 299-312.

Calle, Y., Chou, H. C., Thrasher, A. J. and Jones, G. E. (2004). Wiskott-Aldrich syndrome protein and the cytoskeletal dynamics of dendritic cells. *J Pathol* **204**, 460-9.

Chambers, D. N. and Bretscher, A. (2005). Ezrin mutants affecting dimerization and activation. *Biochemistry* **44**, 3926-32.

Charboneau, A. L., Singh, V., Yu, T. and Newsham, I. F. (2002). Suppression of growth and increased cellular attachment after expression of DAL-1 in MCF-7 breast cancer cells. *Int J Cancer* **100**, 181-8.

Chen, B., Blair, D. G., Plisov, S., Vasiliev, G., Perantoni, A. O., Chen, Q., Athanasiou, M., Wu, J. Y., Oppenheim, J. J. and Yang, D. (2004). Cutting edge: bone morphogenetic protein antagonists Drm/Gremlin and Dan interact with Slits and act as negative regulators of monocyte chemotaxis. *J Immunol* **173**, 5914-7.

Chen, W. T. (1989). Proteolytic activity of specialized surface protrusions formed at rosette contact sites of transformed cells. *J Exp Zool* **251**, 167-85.

Chenna, R., Sugawara, H., Koike, T., Lopez, R., Gibson, T. J., Higgins, D. G. and Thompson, J. D. (2003). Multiple sequence alignment with the Clustal series of programs. *Nucleic Acids Res* **31**, 3497-500.

Chishti, A. H., Kim, A. C., Marfatia, S. M., Lutchman, M., Hanspal, M., Jindal, H., Liu, S. C., Low, P. S., Rouleau, G. A., Mohandas, N. et al. (1998). The FERM domain: a unique module involved in the linkage of cytoplasmic proteins to the membrane. *Trends Biochem Sci* **23**, 281-2.

Chuaqui, R. F., Bonner, R. F., Best, C. J., Gillespie, J. W., Flaig, M. J., Hewitt, S. M., Phillips, J. L., Krizman, D. B., Tangrea, M. A., Ahram, M. et al. (2002). Post-analysis follow-up and validation of microarray experiments. *Nat Genet* **32 Suppl**, 509-14.

Clark, E. A., Golub, T. R., Lander, E. S. and Hynes, R. O. (2000). Genomic analysis of metastasis reveals an essential role for RhoC. *Nature* **406**, 532-5.

Cohen, A. R., Woods, D. F., Marfatia, S. M., Walther, Z., Chishti, A. H., Anderson, J. M. and Wood, D. F. (1998). Human CASK/LIN-2 binds syndecan-2 and protein 4.1 and localizes to the basolateral membrane of epithelial cells. *J Cell Biol* **142**, 129-38.

Conboy, J., Kan, Y. W., Shohet, S. B. and Mohandas, N. (1986). Molecular cloning of protein 4.1, a major structural element of the human erythrocyte membrane skeleton. *Proc Natl Acad Sci U S A* **83**, 9512-6.

Condeelis, J., Singer, R. H. and Segall, J. E. (2005). The great escape: when cancer cells hijack the genes for chemotaxis and motility. *Annu Rev Cell Dev Biol* **21**, 695-718.

Cuppen, E., Gerrits, H., Pepers, B., Wieringa, B. and Hendriks, W. (1998). PDZ motifs in PTP-BL and RIL bind to internal protein segments in the LIM domain protein RIL. *Mol Biol Cell* **9**, 671-83.

Dalton, W. S. and Friend, S. H. (2006). Cancer biomarkers--an invitation to the table. *Science* **312**, 1165-8.

de Reynies, A., Geromin, D., Cayuela, J. M., Petel, F., Dessen, P., Sigaux, F. and Rickman, D. S. (2006). Comparison of the latest commercial short and long oligonucleotide microarray technologies. *BMC Genomics* **7**, 51.

Duggan, D. J., Bittner, M., Chen, Y., Meltzer, P. and Trent, J. M. (1999). Expression profiling using cDNA microarrays. *Nat Genet* **21**, 10-4.

Dunn, G. A., Dobbie, I. M., Monypenny, J., Holt, M. R. and Zicha, D. (2002). Fluorescence localization after photobleaching (FLAP): a new method for studying protein dynamics in living cells. *J Microsc* **205**, 109-12.

Ein-Dor, L., Zuk, O. and Domany, E. (2006). Thousands of samples are needed to generate a robust gene list for predicting outcome in cancer. *Proc Natl Acad Sci U S A* **103**, 5923-8.

Endo, A., Surks, H. K., Mochizuki, S., Mochizuki, N. and Mendelsohn, M. E. (2004). Identification and characterization of zipper-interacting protein kinase as the unique vascular smooth muscle myosin phosphatase-associated kinase. *J Biol Chem* **279**, 42055-61.

Evans, J. G., Correia, I., Krasavina, O., Watson, N. and Matsudaira, P. (2003). Macrophage podosomes assemble at the leading lamella by growth and fragmentation. *J Cell Biol* **161**, 697-705.

Faassen, A. E., Mooradian, D. L., Tranquillo, R. T., Dickinson, R. B., Letourneau, P. C., Oegema, T. R. and McCarthy, J. B. (1993). Cell surface CD44-related chondroitin sulfate proteoglycan is required for transforming growth factor-beta-stimulated mouse melanoma cell motility and invasive behavior on type I collagen. *J Cell Sci* **105 (Pt 2)**, 501-11.

Fidler, I. J. (1973). Selection of successive tumour lines for metastasis. *Nat New Biol* **242**, 148-9.

Fidler, I. J. (2002). Critical determinants of metastasis. *Semin Cancer Biol* **12**, 89-96.

Fidler, I. J. and Ellis, L. M. (1994). The implications of angiogenesis for the biology and therapy of cancer metastasis. *Cell* **79**, 185-8.

Fidler, I. J. and Hart, I. R. (1982). Biological diversity in metastatic neoplasms: origins and implications. *Science* **217**, 998-1003.

Friedl, P. and Wolf, K. (2003). Proteolytic and non-proteolytic migration of tumour cells and leucocytes. *Biochem Soc Symp*, 277-85.

Fu, S. L., Waha, A. and Vogt, P. K. (2000). Identification and characterization of genes upregulated in cells transformed by v-Jun. *Oncogene* **19**, 3537-45.

Fukuda, M. and Mikoshiba, K. (2001). The N-terminal cysteine cluster is essential for membrane targeting of B/K protein. *Biochem J* **360**, 441-8.

Gascard, P., Parra, M. K., Zhao, Z., Calinisan, V. R., Nunomura, W., Rivkees, S. A., Mohandas, N. and Conboy, J. G. (2004). Putative tumor suppressor protein 4.1B is differentially expressed in kidney and brain via alternative promoters and 5' alternative splicing. *Biochim Biophys Acta* **1680**, 71-82.

Gerber, M. A., Bahr, S. M. and Gutmann, D. H. (2006). Protein 4.1B/Differentially Expressed in Adenocarcinoma of the Lung-1 Functions as a Growth Suppressor in Meningioma Cells by Activating Rac1-Dependent c-Jun-NH2-kinase Signaling. *Cancer Res* **66**, 5295-303.

Gildea, J. J., Seraj, M. J., Oxford, G., Harding, M. A., Hampton, G. M., Moskaluk, C. A., Frierson, H. F., Conaway, M. R. and Theodorescu, D. (2002). RhoGDI2 is an invasion and metastasis suppressor gene in human cancer. *Cancer Res* **62**, 6418-23.

Gimm, J. A., An, X., Nunomura, W. and Mohandas, N. (2002). Functional characterization of spectrin-actin-binding domains in 4.1 family of proteins. *Biochemistry* **41**, 7275-82.

Girnita, A., All-Ericsson, C., Economou, M. A., Astrom, K., Axelson, M., Seregard, S., Larsson, O. and Girnita, L. (2006). The insulin-like growth factor-I receptor inhibitor picropodophyllin causes tumor regression and attenuates mechanisms involved in invasion of uveal melanoma cells. *Clin Cancer Res* **12**, 1383-91.

Gould, K. L., Bretscher, A., Esch, F. S. and Hunter, T. (1989). cDNA cloning and sequencing of the protein-tyrosine kinase substrate, ezrin, reveals homology to band 4.1. *Embo J* **8**, 4133-42.

Gutmann, D. H., Donahoe, J., Perry, A., Lemke, N., Gorse, K., Kittiniyom, K., Rempel, S. A., Gutierrez, J. A. and Newsham, I. F. (2000). Loss of DAL-1, a protein 4.1-related tumor suppressor, is an important early event in the pathogenesis of meningiomas. *Hum Mol Genet* **9**, 1495-500.

Gutmann, D. H., Hirbe, A. C., Huang, Z. Y. and Haipek, C. A. (2001). The protein 4.1 tumor suppressor, DAL-1, impairs cell motility, but regulates proliferation in a cell-type-specific fashion. *Neurobiol Dis* **8**, 266-78.

Guvakova, M. A., Adams, J. C. and Boettiger, D. (2002). Functional role of alpha-actinin, PI 3-kinase and MEK1/2 in insulin-like growth factor I receptor kinase regulated motility of human breast carcinoma cells. *J Cell Sci* **115**, 4149-65.

Hahn, W. C. and Weinberg, R. A. (2002). Rules for making human tumor cells. *N Engl J Med* **347**, 1593-603.

Hart, C. A., Brown, M., Bagley, S., Sharrard, M. and Clarke, N. W. (2005). Invasive characteristics of human prostatic epithelial cells: understanding the metastatic process. *Br J Cancer* **92**, 503-12.

Hata, Y., Butz, S. and Sudhof, T. C. (1996). CASK: a novel dlg/PSD95 homolog with an N-terminal calmodulin-dependent protein kinase domain identified by interaction with neurexins. *J Neurosci* **16**, 2488-94.

Hujanen, E. S. and Terranova, V. P. (1985). Migration of tumor cells to organ-derived chemoattractants. *Cancer Res* **45**, 3517-21.

Ilic, D., Furuta, Y., Kanazawa, S., Takeda, N., Sobue, K., Nakatsuji, N., Nomura, S., Fujimoto, J., Okada, M. and Yamamoto, T. (1995). Reduced cell motility and enhanced focal adhesion contact formation in cells from FAK-deficient mice. *Nature* **377**, 539-44.

Jechlinger, M., Sommer, A., Moriggl, R., Seither, P., Kraut, N., Capodiecci, P., Donovan, M., Cordon-Cardo, C., Beug, H. and Grunert, S. (2006). Autocrine PDGFR signaling promotes mammary cancer metastasis. *J Clin Invest* **116**, 1561-1570.

Jiang, W. and Newsham, I. F. (2006). The tumor suppressor DAL-1/4.1B and protein methylation cooperate in inducing apoptosis in MCF-7 breast cancer cells. *Mol Cancer* **5**, 4.

Jiang, W., Roemer, M. E. and Newsham, I. F. (2005). The tumor suppressor DAL-1/4.1B modulates protein arginine N-methyltransferase 5 activity in a substrate-specific manner. *Biochem Biophys Res Commun* **329**, 522-30.

Jonsson, P. F., Cavanna, T., Zicha, D. and Bates, P. A. (2006). Cluster analysis of networks generated through homology: automatic identification of important protein communities involved in cancer metastasis. *BMC Bioinformatics* **7**, 2.

Jordan, M. A. and Wilson, L. (2004). Microtubules as a target for anticancer drugs. *Nat Rev Cancer* **4**, 253-65.

Kankkunen, J. P., Harvima, I. T. and Naukkarinen, A. (1997). Quantitative analysis of tryptase and chymase containing mast cells in benign and malignant breast lesions. *Int J Cancer* **72**, 385-8.

Karpatkin, S. and Pearlstein, E. (1981). Role of platelets in tumor cell metastases. *Ann Intern Med* **95**, 636-41.

Katoh, K., Kano, Y., Amano, M., Onishi, H., Kaibuchi, K. and Fujiwara, K. (2001). Rho-kinase--mediated contraction of isolated stress fibers. *J Cell Biol* **153**, 569-84.

Khanna, C., Khan, J., Nguyen, P., Prehn, J., Caylor, J., Yeung, C., Trepel, J., Meltzer, P. and Helman, L. (2001). Metastasis-associated differences in gene expression in a murine model of osteosarcoma. *Cancer Res* **61**, 3750-9.

Khanna, C., Wan, X., Bose, S., Cassaday, R., Olomu, O., Mendoza, A., Yeung, C., Gorlick, R., Hewitt, S. M. and Helman, L. J. (2004). The membrane-cytoskeleton linker ezrin is necessary for osteosarcoma metastasis. *Nat Med* **10**, 182-6.

Kiewe, P., Bechrakis, N. E., Schmittel, A., Ruf, P., Lindhofer, H., Thiel, E. and Nagorsen, D. (2006). Increased chondroitin sulphate proteoglycan expression (B5 immunoreactivity) in metastases of uveal melanoma. *Ann Oncol* **17**, 1830-4.

Kikuchi, S., Yamada, D., Fukami, T., Maruyama, T., Ito, A., Asamura, H., Matsuno, Y., Onizuka, M. and Murakami, Y. (2006). Hypermethylation of the TSLC1/IGSF4 promoter is associated with tobacco smoking and a poor prognosis in primary nonsmall cell lung carcinoma. *Cancer* **106**, 1751-8.

Kikuchi, S., Yamada, D., Fukami, T., Masuda, M., Sakurai-Yageta, M., Williams, Y. N., Maruyama, T., Asamura, H., Matsuno, Y., Onizuka, M. et al.

(2005). Promoter methylation of DAL-1/4.1B predicts poor prognosis in non-small cell lung cancer. *Clin Cancer Res* **11**, 2954-61.

Kittiniyom, K., Gorse, K. M., Dalbegue, F., Lichy, J. H., Taubenberger, J. K. and Newsham, I. F. (2001). Allelic loss on chromosome band 18p11.3 occurs early and reveals heterogeneity in breast cancer progression. *Breast Cancer Res* **3**, 192-8.

Klijn, J. G., Look, M. P., Portengen, H., Alexieva-Figusch, J., van Putten, W. L. and Foekens, J. A. (1994). The prognostic value of epidermal growth factor receptor (EGF-R) in primary breast cancer: results of a 10 year follow-up study. *Breast Cancer Res Treat* **29**, 73-83.

Kuno, K., Bannai, K., Hakozaki, M., Matsushima, K. and Hirose, K. (2004). The carboxyl-terminal half region of ADAMTS-1 suppresses both tumorigenicity and experimental tumor metastatic potential. *Biochem Biophys Res Commun* **319**, 1327-33.

Kwon, O. J., Gainer, H., Wray, S. and Chin, H. (1996). Identification of a novel protein containing two C2 domains selectively expressed in the rat brain and kidney. *FEBS Lett* **378**, 135-9.

Lacal, P. M., Ruffini, F., Pagani, E. and D'Atri, S. (2005). An autocrine loop directed by the vascular endothelial growth factor promotes invasiveness of human melanoma cells. *Int J Oncol* **27**, 1625-32.

Lankes, W. T. and Furthmayr, H. (1991). Moesin: a member of the protein 4.1-talin-ezrin family of proteins. *Proc Natl Acad Sci U S A* **88**, 8297-301.

Lappalainen, P. and Drubin, D. G. (1997). Cofilin promotes rapid actin filament turnover in vivo. *Nature* **388**, 78-82.

Lee, J. E., Kim, H. J., Bae, J. Y., Kim, S. W., Park, J. S., Shin, H. J., Han, W., Kang, K. S. and Noh, D. Y. (2005). Neogenin expression may be inversely correlated to the tumorigenicity of human breast cancer. *BMC Cancer* **5**, 154.

Li, Y., Tang, Y., Ye, L., Liu, B., Liu, K., Chen, J. and Xue, Q. (2003). Establishment of a hepatocellular carcinoma cell line with unique metastatic characteristics through in vivo selection and screening for metastasis-related genes through cDNA microarray. *J Cancer Res Clin Oncol* **129**, 43-51.

Liu, J., Blackhall, F., Seiden-Long, I., Jurisica, I., Navab, R., Liu, N., Radulovich, N., Wigle, D., Sultan, M., Hu, J. et al. (2004). Modeling of lung cancer by an orthotopically growing H460SM variant cell line reveals novel candidate genes for systemic metastasis. *Oncogene* **23**, 6316-24.

Lockhart, D. J., Dong, H., Byrne, M. C., Follettie, M. T., Gallo, M. V., Chee, M. S., Mittmann, M., Wang, C., Kobayashi, M., Horton, H. et al. (1996). Expression monitoring by hybridization to high-density oligonucleotide arrays. *Nat Biotechnol* **14**, 1675-80.

Lodish, H., Berk, A., Zipursky, S., Matsudaira, P., Baltimore, D., Darnell, J. . (1999). Molecular Cell Biology: W. H. Freeman and Co.

MacDonald, J. A., Borman, M. A., Muranyi, A., Somlyo, A. V., Hartshorne, D. J. and Haystead, T. A. (2001). Identification of the endogenous smooth muscle myosin phosphatase-associated kinase. *Proc Natl Acad Sci U S A* **98**, 2419-24.

Mackay, D. J., Esch, F., Furthmayr, H. and Hall, A. (1997). Rho- and rac-dependent assembly of focal adhesion complexes and actin filaments in permeabilized fibroblasts: an essential role for ezrin/radixin/moesin proteins. *J Cell Biol* **138**, 927-38.

Margalit, O., Eisenbach, L., Amariglio, N., Kaminski, N., Harmelin, A., Pfeffer, R., Shohat, M., Rechavi, G. and Berger, R. (2003). Overexpression of a set of genes, including WISP-1, common to pulmonary metastases of both mouse D122 Lewis lung carcinoma and B16-F10.9 melanoma cell lines. *Br J Cancer* **89**, 314-9.

Marhaba, R. and Zoller, M. (2004). CD44 in cancer progression: adhesion, migration and growth regulation. *J Mol Histol* **35**, 211-31.

Martins, G. G. and Kolega, J. (2006). Endothelial cell protrusion and migration in three-dimensional collagen matrices. *Cell Motil Cytoskeleton* **63**, 101-15.

McBride, A. E. and Silver, P. A. (2001). State of the arg: protein methylation at arginine comes of age. *Cell* **106**, 5-8.

McCarty, J. H., Cook, A. A. and Hynes, R. O. (2005). An interaction between {alpha}v{beta}8 integrin and Band 4.1B via a highly conserved region of the Band 4.1 C-terminal domain. *Proc Natl Acad Sci U S A* **102**, 13479-83.

McClatchey, A. I. and Giovannini, M. (2005). Membrane organization and tumorigenesis--the NF2 tumor suppressor, Merlin. *Genes Dev* **19**, 2265-77.

Michiels, S., Koscielny, S. and Hill, C. (2005). Prediction of cancer outcome with microarrays: a multiple random validation strategy. *Lancet* **365**, 488-92.

Miyata, H. and Hotani, H. (1992). Morphological changes in liposomes caused by polymerization of encapsulated actin and spontaneous formation of actin bundles. *Proc Natl Acad Sci U S A* **89**, 11547-51.

Monypenny, J. (2003). The development of quantitative live cell imaging techniques and their applications in the study of inter-cellular communication and sarcoma cell motility. In *Biochemistry*, vol. PhD. London: University College London.

Muller, A., Homey, B., Soto, H., Ge, N., Catron, D., Buchanan, M. E., McClanahan, T., Murphy, E., Yuan, W., Wagner, S. N. et al. (2001). Involvement of chemokine receptors in breast cancer metastasis. *Nature* **410**, 50-6.

Niggemann, B., Drell, T. L. t., Joseph, J., Weidt, C., Lang, K., Zaenker, K. S. and Entschladen, F. (2004). Tumor cell locomotion: differential dynamics of spontaneous and induced migration in a 3D collagen matrix. *Exp Cell Res* **298**, 178-87.

Pagel, P., Kovac, S., Oesterheld, M., Brauner, B., Dunger-Kaltenbach, I., Frishman, G., Montrone, C., Mark, P., Stumpflen, V., Mewes, H. W. et al. (2005). The MIPS mammalian protein-protein interaction database. *Bioinformatics* **21**, 832-4.

Parra, M., Gascard, P., Walensky, L. D., Gimm, J. A., Blackshaw, S., Chan, N., Takakuwa, Y., Berger, T., Lee, G., Chasis, J. A. et al. (2000). Molecular and functional characterization of protein 4.1B, a novel member of the protein 4.1 family with high level, focal expression in brain. *J Biol Chem* **275**, 3247-55.

Parra, M., Gascard, P., Walensky, L. D., Snyder, S. H., Mohandas, N. and Conboy, J. G. (1998). Cloning and characterization of 4.1G (EPB41L2), a new member of the skeletal protein 4.1 (EPB41) gene family. *Genomics* **49**, 298-306.

Parsons, J. T., Martin, K. H., Slack, J. K., Taylor, J. M. and Weed, S. A. (2000). Focal adhesion kinase: a regulator of focal adhesion dynamics and cell movement. *Oncogene* **19**, 5606-13.

Pawitan, Y., Bjohle, J., Amler, L., Borg, A. L., Egyhazi, S., Hall, P., Han, X., Holmberg, L., Huang, F., Klaar, S. et al. (2005). Gene expression profiling spares early breast cancer patients from adjuvant therapy: derived and validated in two population-based cohorts. *Breast Cancer Res* **7**, R953-64.

Peracaula, R., Cleary, K. R., Lorenzo, J., de Llorens, R. and Frazier, M. L. (2000). Human pancreatic ribonuclease 1: expression and distribution in pancreatic adenocarcinoma. *Cancer* **89**, 1252-8.

Pokorna, E., Jordan, P. W., O'Neill, C. H., Zicha, D., Gilbert, C. S. and Vesely, P. (1994). Actin cytoskeleton and motility in rat sarcoma cell populations with different metastatic potential. *Cell Motil Cytoskeleton* **28**, 25-33.

Pollard, T. D., Blanchoin, L. and Mullins, R. D. (2000). Molecular mechanisms controlling actin filament dynamics in nonmuscle cells. *Annu Rev Biophys Biomol Struct* **29**, 545-76.

Ramaswamy, S., Ross, K. N., Lander, E. S. and Golub, T. R. (2003). A molecular signature of metastasis in primary solid tumors. *Nat Genet* **33**, 49-54.

Rao, J. and Li, N. (2004). Microfilament actin remodeling as a potential target for cancer drug development. *Curr Cancer Drug Targets* **4**, 345-54.

Rew, D. A. (2001). DNA microarray technology in cancer research. *Eur J Surg Oncol* **27**, 504-8.

Ridley, A. J. and Hall, A. (1992). The small GTP-binding protein rho regulates the assembly of focal adhesions and actin stress fibers in response to growth factors. *Cell* **70**, 389-99.

Robb, V. A., Li, W. and Gutmann, D. H. (2004). Disruption of 14-3-3 binding does not impair Protein 4.1B growth suppression. *Oncogene* **23**, 3589-96.

Sahai, E. (2005). Mechanisms of cancer cell invasion. *Curr Opin Genet Dev* **15**, 87-96.

Sahai, E. and Marshall, C. J. (2002). ROCK and Dia have opposing effects on adherens junctions downstream of Rho. *Nat Cell Biol* **4**, 408-15.

Sahai, E., Wyckoff, J., Philippar, U., Segall, J. E., Gertler, F. and Condeelis, J. (2005). Simultaneous imaging of GFP, CFP and collagen in tumors in vivo using multiphoton microscopy. *BMC Biotechnol* **5**, 14.

Salwinski, L., Miller, C. S., Smith, A. J., Pettit, F. K., Bowie, J. U. and Eisenberg, D. (2004). The Database of Interacting Proteins: 2004 update. *Nucleic Acids Res* **32**, D449-51.

Sastry, S. and Burridge, K. (2000). Focal Adhesions: A Nexus for Intracellular Signalling and Cytoskeletal Dynamics. *Experimental Cell Res* **261**, 25-36.

Sato, T., Sakai, T., Noguchi, Y., Takita, M., Hirakawa, S. and Ito, A. (2004). Tumor-stromal cell contact promotes invasion of human uterine cervical carcinoma cells by augmenting the expression and activation of stromal matrix metalloproteinases. *Gynecol Oncol* **92**, 47-56.

Silverman, R. H. (2003). Implications for RNase L in prostate cancer biology. *Biochemistry* **42**, 1805-12.

Singh, V., Miranda, T. B., Jiang, W., Frankel, A., Roemer, M. E., Robb, V. A., Gutmann, D. H., Herschman, H. R., Clarke, S. and Newsham, I. F. (2004). DAL-1/4.1B tumor suppressor interacts with protein arginine N-methyltransferase 3 (PRMT3) and inhibits its ability to methylate substrates in vitro and in vivo. *Oncogene* **23**, 7761-71.

Small, J. V., Stradal, T., Vignal, E. and Rottner, K. (2002). The lamellipodium: where motility begins. *Trends Cell Biol* **12**, 112-20.

Soderling, T. R. (1999). The Ca-calmodulin-dependent protein kinase cascade. *Trends Biochem Sci* **24**, 232-6.

Svitkina, T., Verkhovsky, A. and Borisy, G. (1997). Recent advances in microscopy of cells, tissues and organs: Antonio Delfino Editore S.r.l., Rome.

Svitkina, T. M. and Borisy, G. G. (1999). Progress in protrusion: the tell-tale scar. *Trends Biochem Sci* **24**, 432-6.

Thiery, J. P. (2003). Epithelial-mesenchymal transitions in development and pathologies. *Curr Opin Cell Biol* **15**, 740-6.

Tilney, L. G. (1978). Polymerization of actin. V. A new organelle, the actomere, that initates the assembly of actin filaments in Thyone sperm. *J Cell Biol* **77**, 851-64.

Topol, L. Z., Marx, M., Laugier, D., Bogdanova, N. N., Boubnov, N. V., Clausen, P. A., Calothy, G. and Blair, D. G. (1997). Identification of drm, a novel gene whose expression is suppressed in transformed cells and which can inhibit growth of normal but not transformed cells in culture. *Mol Cell Biol* **17**, 4801-10.

Tran, Y., Benbatoul, K., Gorse, K., Rempel, S., Futreal, A., Green, M. and Newsham, I. (1998). Novel regions of allelic deletion on chromosome 18p in tumors of the lung, brain and breast. *Oncogene* **17**, 3499-505.

Tran, Y. K., Bogler, O., Gorse, K. M., Wieland, I., Green, M. R. and Newsham, I. F. (1999). A novel member of the NF2/ERM/4.1 superfamily with growth suppressing properties in lung cancer. *Cancer Res* **59**, 35-43.

Tsukita, S., Hieda, Y. and Tsukita, S. (1989). A new 82-kD barbed end-capping protein (radixin) localized in the cell-to-cell adherens junction: purification and characterization. *J Cell Biol* **108**, 2369-82.

Vallenius, T., Scharm, B., Vesikansa, A., Luukko, K., Schafer, R. and Makela, T. P. (2004). The PDZ-LIM protein RIL modulates actin stress fiber turnover and enhances the association of alpha-actinin with F-actin. *Exp Cell Res* **293**, 117-28.

van 't Veer, L. J., Dai, H., van de Vijver, M. J., He, Y. D., Hart, A. A., Mao, M., Peterse, H. L., van der Kooy, K., Marton, M. J., Witteveen, A. T. et al. (2002). Gene expression profiling predicts clinical outcome of breast cancer. *Nature* **415**, 530-6.

van de Vijver, M. J., He, Y. D., van't Veer, L. J., Dai, H., Hart, A. A., Voskuil, D. W., Schreiber, G. J., Peterse, J. L., Roberts, C., Marton, M. J. et al. (2002). A gene-expression signature as a predictor of survival in breast cancer. *N Engl J Med* **347**, 1999-2009.

Vesely. (1989). Patterns of in vitro locomotor behaviour characterising cells of spontaneously metastasising rat sarcomas. *Highlights Modern Biochem* **2**, 973-981.

Vesely, P., Chaloupkova, A., Urbanec, P., Urbancova, H., Bohac, L., Krchnakova, E., Franc, F., Sprincl, L., Vousden, K., Moss, R. et al. (1987). Patterns of in vitro behaviour characterizing cells of spontaneously metastasizing K2M rat sarcoma. *Folia Biol (Praha)* **33**, 307-24.

Vesely, P. and Weiss, R. A. (1973). Cell locomotion and contact inhibition of normal and neoplastic rat cells. *Int J Cancer* **11**, 64-76.

Walensky, L. D., Blackshaw, S., Liao, D., Watkins, C. C., Weier, H. U., Parra, M., Huganir, R. L., Conboy, J. G., Mohandas, N. and Snyder, S. H. (1999). A novel neuron-enriched homolog of the erythrocyte membrane cytoskeletal protein 4.1. *J Neurosci* **19**, 6457-67.

Wang, Q., Lu, J., Yang, C., Wang, X., Cheng, L., Hu, G., Sun, Y., Zhang, X., Wu, M. and Liu, Z. (2002a). CASK and its target gene Reelin were co-upregulated in human esophageal carcinoma. *Cancer Lett* **179**, 71-7.

Wang, W., Wyckoff, J. B., Frohlich, V. C., Oleynikov, Y., Huttelmaier, S., Zavadil, J., Cermak, L., Bottinger, E. P., Singer, R. H., White, J. G. et al. (2002b). Single cell behavior in metastatic primary mammary tumors correlated with gene expression patterns revealed by molecular profiling. *Cancer Res* **62**, 6278-88.

Wang, W., Wyckoff, J. B., Wang, Y., Bottinger, E. P., Segall, J. E. and Condeelis, J. S. (2003). Gene expression analysis on small numbers of invasive cells collected by chemotaxis from primary mammary tumors of the mouse. *BMC Biotechnol* **3**, 13.

Wang, Y., Klijn, J. G., Zhang, Y., Sieuwerts, A. M., Look, M. P., Yang, F., Talantov, D., Timmermans, M., Meijer-van Gelder, M. E., Yu, J. et al. (2005). Gene-expression profiles to predict distant metastasis of lymph-node-negative primary breast cancer. *Lancet* **365**, 671-9.

Waterman-Storer, C. M. and Salmon, E. D. (1997). Microtubule dynamics: treadmilling comes around again. *Curr Biol* **7**, R369-72.

Wyckoff, J. B., Jones, J. G., Condeelis, J. S. and Segall, J. E. (2000). A critical step in metastasis: in vivo analysis of intravasation at the primary tumor. *Cancer Res* **60**, 2504-11.

Yageta, M., Kuramochi, M., Masuda, M., Fukami, T., Fukuhara, H., Maruyama, T., Shibuya, M. and Murakami, Y. (2002). Direct association of TSLC1 and DAL-1, two distinct tumor suppressor proteins in lung cancer. *Cancer Res* **62**, 5129-33.

Yang, F., Foekens, J. A., Yu, J., Sieuwerts, A. M., Timmermans, M., Klijn, J. G., Atkins, D., Wang, Y. and Jiang, Y. (2006). Laser microdissection and microarray analysis of breast tumors reveal ER-alpha related genes and pathways. *Oncogene* **25**, 1413-9.

Yang, J., Mani, S. A., Donaher, J. L., Ramaswamy, S., Itzykson, R. A., Come, C., Savagner, P., Gitelman, I., Richardson, A. and Weinberg, R. A. (2004). Twist, a master regulator of morphogenesis, plays an essential role in tumor metastasis. *Cell* **117**, 927-39.

Yi, C., McCarty, J. H., Troutman, S. A., Eckman, M. S., Bronson, R. T. and Kissil, J. L. (2005). Loss of the putative tumor suppressor band 4.1B/Dal1 gene is dispensable for normal development and does not predispose to cancer. *Mol Cell Biol* **25**, 10052-9.

Yu, T., Robb, V. A., Singh, V., Gutmann, D. H. and Newsham, I. F. (2002). The 4.1/ezrin/radixin/moesin domain of the DAL-1/Protein 4.1B tumour suppressor interacts with 14-3-3 proteins. *Biochem J* **365**, 783-9.

Zicha, D., Dobbie, I. M., Holt, M. R., Monypenny, J., Soong, D. Y., Gray, C. and Dunn, G. A. (2003). Rapid actin transport during cell protrusion. *Science* **300**, 142-5.

Zicha, D. and Dunn, G. A. (1995). Are growth factors chemotactic agents? *Exp Cell Res* **221**, 526-9.

Zicha, D., Dunn, G. A. and Brown, A. F. (1991). A new direct-viewing chemotaxis chamber. *J Cell Sci* **99 (Pt 4)**, 769-75.

Appendices

Appendix 1: Algorithms used for the processing of cell tracks.

SPECIFY DIRECTORY WITH *.cel (Kinetic Imaging) or *.log (Universal Imaging) FILES AND GROUPS

Files grouped in subdirectories 01_group1, 02_group2, ... also optionally containing *.txt files

*Optional *.txt files can provide following parameters*:
xyResolution = {1.27,1.27} (*in µm/pixel or pixels/Cm (default)*)
resolutionUnit="µm/pixel"
imageWidth=640
imageHeight=512
lapseInterval = 5 (*min*)
rotationAngle = 0 (*degrees*)

Universal Imaging log files from automatic tracking have format: "Object #","Frame #","X","Y","Distance" with log column titles ON

```
dirPath = "Z:\\chemotaxis_groups\\16_niceA297nbFilm\\";
allFiles = FileNames["*.cel", dirPath, Infinity];
groupFilenames = Split[allFiles, dirMatch];
groupNames = Table[StringTake[groupFilenames[[i, 1]],
    {StringPosition[groupFilenames[[i, 1]], "\\"][[-2, 1]] + 4,
     StringPosition[groupFilenames[[i, 1]], "\\"][[-1, 1]] - 1}], {i, Length[groupFilenames]}];
TableForm[Table[{groupNames[[i]], Length[groupFilenames[[i]]]}, {i, Length[groupNames]}], TableHeadings → {Automatic, {"Group", "N"}}]
```

	Group	N
1	niceA297nbFilm	1

SELECTING SUBSET OF GROUPS: False or True If True type in the groups that you'd like to include in the "{...}"

selectGroupSubset = False

```
False
```

```
groupsIncluded = Range[Length[groupFilenames]] (*default to all data*)
```

```
{1}
```

```
If[selectGroupSubset, groupsIncluded = {3, 4}]; TableForm[Table[{groupNames[[groupsIncluded[[i]]]], Length[groupFilenames[[groupsIncluded[[i]]]]]}, {i, Length[groupsIncluded]}],
 TableHeadings -> {Automatic, {"Group", "N"}}]
```

	Group	N
1	niceA297nbFilm	1

Mathematica® routine for the evaluation of cell motility

This reproduction of a Mathematica® Notebook shows the routines used for evaluation of cell motility, from import of the cell tracks to generation of trajectories and production of circular histograms

LOAD TRACK DATA (OPTIONALLY SEPARATE DATA for CNT and TRE with pathological FLAG: False or True)

separatePathologicalCells = False

False

- **OPTIONAL CHOICE OF pathological-> True or False (*replace All*)**

```
If[! separatePathologicalCells,
 trackData = Table[loadTracks[groupFilenames[[groupsIncluded[[groupIndex]]]][[movieIndex]], pathological → All], {groupIndex, Length[groupsIncluded]},
   {movieIndex, Length[groupFilenames[[groupsIncluded[[groupIndex]]]]]}];
 Dimensions /@ trackData]
```

```
Z:\chemotaxis_groups\16_nicoA297nbFilm\030814c.cel: xyResolution = 7874.02 7874.02 μm/pixel lapseInterval = 5
```

```
{{1, 2764, 3}}
```

```
If[separatePathologicalCells,
 trackDataCNT = Table[loadTracks[groupFilenames[[groupsIncluded[[groupIndex]]]][[movieIndex]], pathological → False], {groupIndex, Length[groupsIncluded]},
   {movieIndex, Length[groupFilenames[[groupsIncluded[[groupIndex]]]]]}];
 trackDataTRE = Table[loadTracks[groupFilenames[[groupsIncluded[[groupIndex]]]][[movieIndex]], pathological → True], {groupIndex, Length[groupsIncluded]},
   {movieIndex, Length[groupFilenames[[groupsIncluded[[groupIndex]]]]]}]; trackData = Join[trackDataCNT, trackDataTRE];
 groupNames = Join[StringJoin[#, "_CNT"] & /@ groupNames, StringJoin[#, "_TRE"] & /@ groupNames];
 groupFilenames = Join[groupFilenames, groupFilenames]; TableForm[Table[{groupNames[[groupsIncluded[[i]]]], Length[groupFilenames[[groupsIncluded[[i]]]]]}, {i, Length[groupsIncluded]}],
  TableHeadings -> {Automatic, {"Group", "N"}}]]
```

- **RECHOOSE GROUPS WITH MIXTURE OF CONTROL (CNT) AND TREATED (PATHOLOGICAL - TRE) CELLS**

selectSeparatedGroupSubset = False

False

```
If[separatePathologicalCells, groupsIncluded = Range[Length[groupFilenames]]]
```

```
If[separatePathologicalCells && selectSeparatedGroupSubset, groupsIncluded = {1, 2, 3}]
```

CHOOSE RANGE OF FRAMES INCLUDED IN CALCULATIONS

changeFrameRange = False

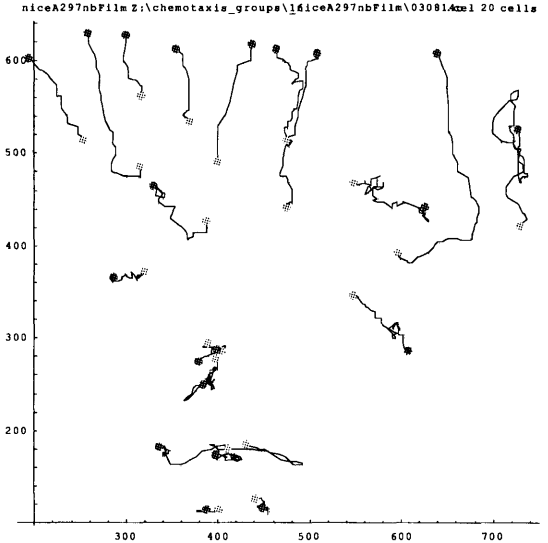
False

```
If[changeFrameRange,
 minFrameNumber = 1;
 maxFrameNumber = 1000;
 trackData = Table[Cases[trackData[[groupIndex, movieIndex]], {c_, f_, {x_, y_, h_}} /; minFrameNumber <= f <= maxFrameNumber], {groupIndex, Length[groupsIncluded]},
   {movieIndex, Length[groupFilenames[[groupsIncluded[[groupIndex]]]]]}];
 Dimensions /@ trackData
]
```

TRACK PLOTS 1

```
Do[localCells = partitionByRuns[Sort[trackData[[groupIndex, movieIndex]]]];
 Show[Graphics[{Line /@ (localCells /. {c_Integer, f_, {ix_, iy_, h_}} → {ix, iy}), RGBColor[0, 1, 0], PointSize[.02],
    (Point /@ Take[#1, 1] &) /@ (localCells /. {c_Integer, f_, {ix_, iy_, h_}} → {ix, iy}), RGBColor[1, 0, 0], PointSize[.02],
    (Point /@ Take[#1, -1] &) /@ (localCells /. {c_Integer, f_, {ix_, iy_, h_}} → {ix, iy})}], AspectRatio → Automatic, Axes → True, PlotRange → All,
  PlotLabel → groupNames[[groupsIncluded[[groupIndex]]]] <> "  " <> groupFilenames[[groupsIncluded[[groupIndex]], movieIndex]] <> " " <> ToString[Length[localCells]] <> " cells"],
 {groupIndex, Length[groupsIncluded]}, {movieIndex, Length[groupFilenames[[groupsIncluded[[groupIndex]]]]]}]
```

```
niceA297nbFilm Z:\chemotaxis_groups\1diceA297nbFilm\030814tm1 20 cells
```



TRACK PLOTS 2

```
tracksFromOrigin = Table[getTracksFromOrigin[trackData[[groupIndex, movieIndex]]], {groupIndex, Length[groupsIncluded]}, {movieIndex, Length[groupFilenames[[groupsIncluded[[groupIndex]]]]]}];
Length /@ tracksFromOrigin
```

```
{1}
```

```
commonPlotRange = {{Min[tracksFromOrigin], Max[tracksFromOrigin]}, {Min[tracksFromOrigin], Max[tracksFromOrigin]}}
```

```
{{-74.93, 214.63}, {-74.93, 214.63}}
```

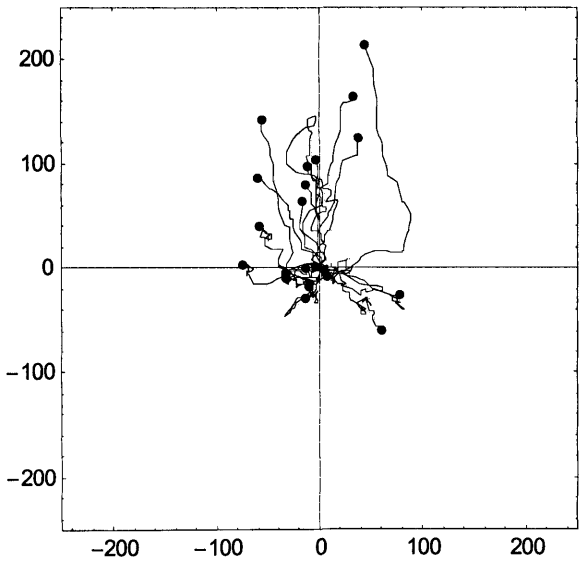
specifyCommonPlotRange = True

```
True
```

```
If[specifyCommonPlotRange, commonPlotRange = {{-250, 250}, {-250, 250}}]
```

```
{{-250, 250}, {-250, 250}}
```

```
Do[
 p[i] =
  Show[
   Graphics[{Text[groupNames[[groupsIncluded[[i]]]] <> "  (N = " <> ToString[Length[Flatten[tracksFromOrigin[[i]], 1]]] <> ")", {0, 275},
     TextStyle → {FontFamily → "Arial", FontSize → 16, FontWeight → "Bold"}], PointSize[.02],
Line /@ Flatten[tracksFromOrigin[[i]], 1], Point[#[[-1]]] & /@ Flatten[tracksFromOrigin[[i]], 1]}], AspectRatio → Automatic, Frame -> True, PlotRange → commonPlotRange,
   GridLines → {{0}, {0}}, TextStyle → {FontFamily → "Arial", FontSize → 16, FontWeight → "Plain"}, PlotRegion → {{0, 1}, {0, 0.90}}], {i, Length[groupsIncluded]}]
```



SECTION 6: RUN = ALWAYS

SPEED ANALYSIS

IMPORTANT: If lapseInterval not specified in Protocol.txt file, it will default to 1 min

```
lapseInterval
```

```
5
```

specifyLapseInterval = False

```
False
```

```
If[specifyLapseInterval, lapseInterval = 1]
```

numberOfLapseIntervalsForSpeedCalculation <1|2|3|4|5...> AND speedUnits <"μm/min"|"μm/hour"|"μm/s"> AND runLength

```
speedData = Table[calculateSpeed[trackData[[groupIndex, movieIndex]], numberOfLapseIntervalsForSpeedCalculation → 1, speedUnits -> "μm/hour", overlap → False, timeStamps → True],
  {groupIndex, Length[groupsIncluded]}, {movieIndex, Length[groupFilenames[[groupsIncluded[[groupIndex]]]]]}];
Length /@ speedData
```

```
{1}
```

Group speed plots

```
commonSpeedPlotRange = {0.0, Max[speedData /. dataPattern → d]}
```

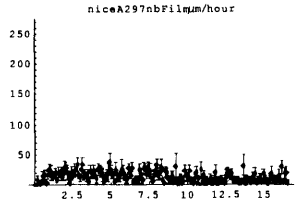
```
{0., 274.743}
```

specifyCommonSpeedPlotRange = False

```
False
```

```
If[specifyCommonSpeedPlotRange, commonSpeedPlotRange = {0, 1000}]
```

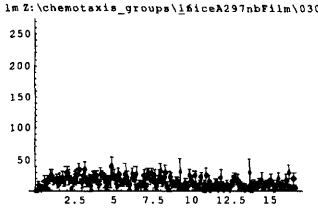
```
Do[MultipleListPlot[splitData=
    Split[Sort[
        Flatten[Flatten[speedData[[groupIndex]],1],
            1]/.dataPattern→{t,d}],#1[[1]]=#2[[1]]&];
plotData=Table[data=Transpose[splitData[[i]]];
    {Mean[data[[1]]],Mean[data[[2]]],
      ErrorBar[StandardErrorOfSampleMean[data[[2]]]]},{i,Length[splitData]}],PlotRange->commonSpeedPlotRange,PlotLabel→
    groupNames[[groupsIncluded[[groupIndex]]]]<>" "<>
      speedUnit],{groupIndex,Length[groupsIncluded]}]
```



Individual *.cel file speed plots

```
Do[MultipleListPlot[splitData=
    Split[Sort[
        Flatten[speedData[[groupIndex,movieIndex]],
            1]/.dataPattern→{t,d}],#1[[1]]=#2[[1]]&];
plotData=Table[data=Transpose[splitData[[i]]];
    {Mean[data[[1]]],Mean[data[[2]]],
      ErrorBar[StandardErrorOfSampleMean[data[[2]]]]},{i,Length[splitData]}],PlotRange->commonSpeedPlotRange,PlotLabel→
    groupNames[[groupsIncluded[[groupIndex]]]]<>"  "<>
      groupFilenames[[groupsIncluded[[groupIndex]],movieIndex]]<>" "<>
      speedUnit],{groupIndex,Length[groupsIncluded]},{movieIndex,
    Length[groupFilenames[[groupsIncluded[[groupIndex]]]]]}]
```

lm Z:\chemotaxis_groups\li6iceA297nbFilm\0308



SPEED SUMMARY

analyseDataSubset = False

False

```
If[analyseDataSubset, negativeConditionPositions = Position[speedData, dataPattern /; d < 0.005];
 dataSubset = Delete[speedData, negativeConditionPositions];
 emptyPos = Position[dataSubset, {}];
 dataSubset = Delete[dataSubset, emptyPos], dataSubset = speedData];
Dimensions[dataSubset]
```

{1, 1, 20}

```
parameterData = dataSubset /. dataPattern → d;
Dimensions[parameterData]
```

{1, 1, 20}

```
referenceGroup = 1
```

1

```
Length /@ parameterData
```

{1}

```
If[MemberQ[Length /@ parameterData, 1], Do[Print[groupNames[[groupsIncluded[[referenceGroup]]]], " vs. ", groupNames[[groupsIncluded[[i]]]]];
  anovaResults[i] = shortANOVA[{parameterData[[referenceGroup, 1]], parameterData[[i, 1]]}], {i, 1, Length[groupsIncluded]}],
 Do[Print[groupNames[[groupsIncluded[[referenceGroup]]]], " vs. ", groupNames[[groupsIncluded[[i]]]]]; anovaResults[i] = shortANOVA[{parameterData[[referenceGroup]], parameterData[[i]]}],
  {i, 1, Length[groupsIncluded]}]]
```

niceA297nbFilm vs. niceA297nbFilm

	Var	DoF1	DoF2	F-ratio	P-value	Signif
L2	-4.88975	1	37.0888	0.		n.s.
L1	76.9331	38	5448.	14.3733	<0.00001	***
L0	781.083	5448				

```
anovaResults[referenceGroup] = {{}, {""}}
```

{{}, {}}

```
meanData = Table[Mean /@ Flatten[parameterData[[groupIndex]], 1], {groupIndex, Length[groupsIncluded]}];
Dimensions[meanData]
```

{1, 20}

```
TableForm[Table[{groupNames[[groupsIncluded[[groupIndex]]]], Length[parameterData[[groupIndex]]], Length[Flatten[parameterData[[groupIndex]], 1]],
   Length[Flatten[parameterData[[groupIndex]], 2]], Mean[meanData[[groupIndex]]], StandardErrorOfSampleMean[meanData[[groupIndex]]], Min[meanData[[groupIndex]]],
   Max[meanData[[groupIndex]]], Median[meanData[[groupIndex]]], anovaResults[groupIndex][[2, -1]]}, {groupIndex, Length[groupsIncluded]}},
 TableHeadings → {Automatic, {"Group", "Movies", "Cells", "Displacements", "Mean", "SEM", "Min", "Max", "Median", "ANOVA"}}]
```

	Group	Movies	Cells	Displacements	Mean	SEM	Min	Max	Median	ANOVA
1	niceA297nbFilm	1	20	2744	17.1537	2.33244	2.79314	39.863	15.4905	

```
speedAndDirectionData = Delete[speedAndDirectionData, Position[speedAndDirectionData, {}]]; Length /@ speedAndDirectionData
```

```
{1}
```

analyseDataSubset = False

```
False
```

```
If[analyseDataSubset, negativeConditionPositions = Position[speedAndDirectionData, dataPattern /; d < 0.005];
 dataSubset = Delete[speedAndDirectionData, negativeConditionPositions];
 emptyPos = Position[dataSubset, {}];
 dataSubset = Delete[dataSubset, emptyPos], dataSubset = speedAndDirectionData];
Dimensions[dataSubset]
```

```
{1, 1, 20}
```

```
parameterData = dataSubset /. dataPattern → p;
Dimensions[parameterData]
```

```
{1, 1, 20}
```

```
referenceGroup = 1
```

```
1
```

```
If[MemberQ[Length /@ parameterData, 1], Do[Print[groupNames[[groupsIncluded[[referenceGroup]]]], " vs. ", groupNames[[groupsIncluded[[i]]]]];
  anovaResults[i] = shortANOVA[{parameterData[[referenceGroup, 1]], parameterData[[i, 1]]}], {i, 1, Length[groupsIncluded]}],
 Do[Print[groupNames[[groupsIncluded[[referenceGroup]]]], " vs. ", groupNames[[groupsIncluded[[i]]]]]; anovaResults[i] = shortANOVA[{parameterData[[referenceGroup]], parameterData[[i]]}],
  {i, 1, Length[groupsIncluded]}]]
```

```
niceA297nbFilm vs. niceA297nbFilm
```

	Var	DoF1	DoF2	F-ratio	P-value	Signif
L2	-0.00116641	1	30.506	0.		n.s.
L1	0.00940437	38	398.	1.77963	0.00390489	**
L0	0.130611	398				

```
anovaResults[referenceGroup] = {{}, {""}}
```

```
{{}, {}}
```

```
meanData = Table[Mean /@ Flatten[parameterData[[groupIndex]], 1], {groupIndex, Length[groupsIncluded]}];
Length /@ meanData
```

```
{20}
```

```
TableForm[Table[{groupNames[[groupsIncluded[[groupIndex]]]], Length[parameterData[[groupIndex]]], Length[Flatten[parameterData[[groupIndex]], 1]],
   Length[Flatten[parameterData[[groupIndex]], 2]], Mean[meanData[[groupIndex]]], StandardErrorOfSampleMean[meanData[[groupIndex]]], Min[meanData[[groupIndex]]],
   Max[meanData[[groupIndex]]], Median[meanData[[groupIndex]]], anovaResults[groupIndex][[2, -1]]}, {groupIndex, Length[groupsIncluded]}},
 TableHeadings → {Automatic, {"Group", "Movies", "Particles", "Displacements", "Mean", "SEM", "Min", "Max", "Median", "ANOVA"}}]
```

	Group	Movies	Particles	Displacements	Mean	SEM	Min	Max	Median	ANOVA
1	niceA297nbFilm	1	20	219	0.736733	0.0357972	0.471095	0.956776	0.753616	

CHOICE OF % OF CELLS TO REACH HORIZON OR HORIZON

```
anglesDyAndSqdist = Table[data = tracksFromOrigin[[groupIndex, movieIndex]];(*horizon=findHorizon[data,50]*); horizon = 70;
   getAnglesDyAndSqdist[data, horizon], {groupIndex, Length[groupsIncluded]},
   {movieIndex, Length[groupFilenames[[groupsIncluded[[groupIndex]]]]]}];
Length /@ anglesDyAndSqdist[[1]]
```

{12}

```
Position[anglesDyAndSqdist, {}]
```

{}

```
anglesDyAndSqdist = Delete[anglesDyAndSqdist, Position[anglesDyAndSqdist, {}]];
```

```
TableForm[Table[data = Mean /@anglesDyAndSqdist[[i]] /. {theta_Real, dy_, sqdist_} :> dy/Sqrt[sqdist];
  {groupNames[[groupsIncluded[[i]]]], Mean[data], StandardErrorOfSampleMean[data]}, {i, Length[groupsIncluded]}]]
```

niceA297nbFilm	0.53687	0

```
referenceGroup = 1
```

1

```
anglesDyAndSqdist[[1]]
```

{{{0.153945, 73.66, 5556.44}, {1.92957, -26.67, 5769.34}, {-0.104877, 72.39, 5298.38}, {0.206683, 78.74, 6472.57}, {-0.495133, 63.5, 5208.05}, {-0.551655, 66.04, 6012.89},
  {0.0535203, 71.12, 5072.57}, {0., 71.12, 5058.05}, {1.35213, 15.24, 4935.47}, {-0.977794, 39.37, 4962.89}, {2.30496, -46.99, 4919.35}, {-1.66146, -6.35, 4919.35}}}

```
{anglesDyAndSqdist[[referenceGroup]], anglesDyAndSqdist[[2]]} /. {theta_Real, dy_, sqdist_} → dy/Sqrt[sqdist]
```

{{{0.988174, -0.351123, 0.994505, 0.978717, 0.879905, 0.851658, 0.998568, 1., 0.21693, 0.558853, -0.669965, -0.0905357}},
 {{{0.988174, -0.351123, 0.994505, 0.978717, 0.879905, 0.851658, 0.998568, 1., 0.21693, 0.558853, -0.669965, -0.0905357}}}[[2]]}

```
Do[Print[groupNames[[groupsIncluded[[referenceGroup]]]] , " vs. ", groupNames[[groupsIncluded[[i]]]]];
 anovaResults1[i] =
  shortANOVA[Flatten[{anglesDyAndSqdist[[referenceGroup]], anglesDyAndSqdist[[i]]} /. {theta_Real, dy_, sqdist_} → dy/Sqrt[sqdist]], 1]
 {i, 1, Length[groupsIncluded]}]
```

niceA297nbFilm vs. niceA297nbFilm

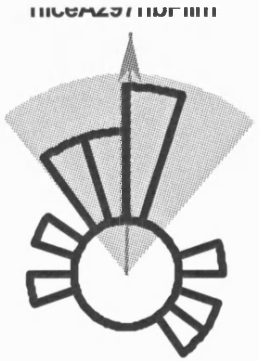
	Var	DoF1	DoF2	F-ratio	P-value	Signif
L0	0.347269	23				

```
anovaResults1[3] = {{}, {""}}
```

{{}, {}}

```
Do[angles = Flatten[anglesDyAndSqdist[[i]] /. {theta_, dy_, sqdist_} → theta];
  taxisReport[groupNames[[groupsIncluded[[i]]]], angles, Length[Flatten[tracksFromOrigin[[i]], 1]]];
  bins1 = BinCounts[angles, {N[-Pi], N[Pi], N[Pi / 10]}];
  p[i] =
   Show[
    {Graphics[{Text[groupNames[[groupsIncluded[[i]]]], {0, 20}, TextStyle → {FontFamily → "Arial", FontSize → 22, FontWeight → "Bold"}],
       Text[anovaResults1[i][[2, -1]] <> "  (N = " <> ToString[Length[angles]] <> ")", {0, -12},
        TextStyle → {FontFamily → "Arial", FontSize → 18, FontWeight → "Bold"}]}], taxisQ[angles], circularHistogram[bins1]},
    AspectRatio → Automatic, PlotRange → {{-15, 15}, {-10, 20}}, PlotRegion → {{0, 1}, {0.1, 0.90}}], {i, 1, Length[groupsIncluded]}];
```

Treatment	niceA297nbFilm
Ntot (total cell tracks) =	20
N (cell tracks reaching horizon) =	12
Mean Direction =	1.24265 degrees
95% Confidence Interval =	-38.5006 to 40.9858 degrees
probability of uniform directions =	0.0308839
kappa =	3.36782



ROUTINES

```
<< "V:\\dan\\mprog\\speedAndChemotaxisAnalysisMaster.m"
```

Appendix 2: Algorithms used for the processing of microarray data.

Mathematica® routine for the analysis of microarray data

This reproduction of a Mathematica® Notebook shows the routines used for the microarray analysis, from import of the data to generation of a candidate gene list to production of graphs of gene expression.


```
annot = Import["D:\\microArray\\RAE230A_annot.csv", "CSV"];
Dimensions[annot]
```

```
{15924, 42}
```

```
blastx = Import["D:\\microArray\\RAE230A_blastx.csv", "CSV"];
Dimensions[blastx]
```

```
{56559, 6}
```

```
homologues = Import["D:\\microArray\\Annotation_1042.csv", "CSV"];
Dimensions[homologues]
```

```
{2223, 7}
```

```
fpData = Get["V:\\dan\\tamara\\fpData2004mANOVA.dat"];
Dimensions[fpData]
```

```
{11859, 11}
```

▪ Import and join unscaled data

```
data1 = Import["D:\\microArray\\unscaled.csv", "CSV"];
Dimensions[data1]
```

```
{15924, 47}
```

```
data2 = Import["D:\\microArray\\unscaled2nd.csv", "CSV"];
Dimensions[data2]
```

```
{15924, 11}
```

```
data3 = Import["D:\\microArray\\unscaled3rd.csv", "CSV"];
Dimensions[data2]
```

```
{15924, 11}
```

```
Dimensions[Drop[data1, 1]]
```

```
{15923, 47}
```

```
Dimensions[Drop[data2, 1]]
```

```
{15923, 11}
```

```
Dimensions[Drop[data3, 1]]
```

```
{15923, 92}
```

```
Transpose[Drop[data1, 1]][[1]] == Transpose[Drop[data2, 1]][[1]]
```

```
True
```

```
Transpose[Drop[data1, 1]][[1]] == Transpose[Drop[data3, 1]][[1]]
```

```
True
```

```
data = Transpose[Join[Drop[Transpose[Drop[data1, 1]], -1], Drop[Drop[Transpose[Drop[data2, 1]], 1], -1], Drop[Transpose[Drop[data3, 1]], 1]]];
Dimensions[data]
```

```
{15923, 146}
```

```
data1[[1]]
```

```
{, 0803_DZ2_uR_A297Nb_5_r1_Signal, 0803_DZ2_uR_A297Nb_5_r1_Detection, 0803_DZ2_uR_A297Nb_5_r1_Detection p-value, 0803_DZ3_uR_A297Nb_9_r1_Signal,
 0803_DZ3_uR_A297Nb_9_r1_Detection, 0803_DZ3_uR_A297Nb_9_r1_Detection p-value, 0803_DZ4_uR_K2_1_r1_Signal, 0803_DZ4_uR_K2_1_r1_Detection,
 0803_DZ4_uR_K2_1_r1_Detection p-value, 0803_DZ5_uR_K2_5_r1_Signal, 0803_DZ5_uR_K2_5_r1_Detection, 0803_DZ5_uR_K2_5_r1_Detection p-value, 0803_DZ6_uR_K2_9_r1_Signal,
 0803_DZ6_uR_K2_9_r1_Detection, 0803_DZ6_uR_K2_9_r1_Detection p-value, 0903_DZ7_uR_A297Nb_2_r2_Signal, 0903_DZ7_uR_A297Nb_2_r2_Detection,
 0903_DZ7_uR_A297Nb_2_r2_Detection p-value, 0903_DZ8_uR_A297Nb_7_r2_Signal, 0903_DZ8_uR_A297Nb_7_r2_Detection, 0903_DZ8_uR_A297Nb_7_r2_Detection p-value,
 0903_DZ9_uR_A297Nb_10_r2_Signal, 0903_DZ9_uR_A297Nb_10_r2_Detection, 0903_DZ9_uR_A297Nb_10_r2_Detection p-value, 0903_DZ10_uR_K2_2_r2_Signal,
 0903_DZ10_uR_K2_2_r2_Detection, 0903_DZ10_uR_K2_2_r2_Detection p-value, 0803_DZ11_uR_K2_6_r2_Signal, 0803_DZ11_uR_K2_6_r2_Detection,
 0803_DZ11_uR_K2_6_r2_Detection p-value, 0803_DZ12_uR_K2_10_r2_Signal, 0803_DZ12_uR_K2_10_r2_Detection, 0803_DZ12_uR_K2_10_r2_Detection p-value,
 0903_DZ13_uR_A297Nb_1_r1_Signal, 0903_DZ13_uR_A297Nb_1_r1_Detection, 0903_DZ13_uR_A297Nb_1_r1_Detection p-value, 0903_DZ14_uR_K2_3_r3_Signal,
 0903_DZ14_uR_K2_3_r3_Detection, 0903_DZ14_uR_K2_3_r3_Detection p-value, 0903_DZ15_uR_K2_7_r3_Signal, 0903_DZ15_uR_K2_7_r3_Detection,
 0903_DZ15_uR_K2_7_r3_Detection p-value, 0903_DZ16_uR_K2_11_r3_Signal, 0903_DZ16_uR_K2_11_r3_Detection, 0903_DZ16_uR_K2_11_r3_Detection p-value, Descriptions}
```

- **Derive data pattern: 1-4 quiescent, 5-8 early, 9-12 late**

```
columnNames = StringDrop[#, StringPosition[#, "_"][[3, 1]]] &/@ Drop[Drop[data1[[1]], -1], 1];
dataPattern1 =
  StringReplace[#, {"A297Nb" → "a", "K2" -> "k", "_1_" -> "c", "_2_" -> "c", "_3_" -> "c", "_4_" -> "c", "_5_" -> "e", "_6_" -> "e", "_7_" -> "e",
      "_8_" -> "e", "_9_" -> "l", "_10_" -> "l", "_11_" -> "l", "_12_" -> "l", "r" -> "", "_Signal" -> "s_", "_Detection p-value" -> "p_", "_Detection" -> "d_"}] &/@
   columnNames;
dataPattern1 = ToExpression[Prepend[dataPattern1, "identifier_"]]
```

```
{identifier_, ae1s_, ae1d_, ae1p_, al1s_, al1d_, al1p_, kc1s_, kc1d_, kc1p_, ke1s_, ke1d_, ke1p_, kl1s_, kl1d_, kl1p_, ac2s_, ac2d_, ac2p_, ae2s_, ae2d_, ae2p_, al2s_,
 al2d_, al2p_, kc2s_, kc2d_, kc2p_, ke2s_, ke2d_, ke2p_, kl2s_, kl2d_, kl2p_, ac1s_, ac1d_, ac1p_, kc3s_, kc3d_, kc3p_, ke3s_, ke3d_, ke3p_, kl3s_, kl3d_, kl3p_}
```

```
columnNames = StringDrop[#, StringPosition[#, "_"][[3, 1]]] &/@ Drop[Drop[data2[[1]], -1], 1];
dataPattern2 =
  StringReplace[#, {"A297Nb" → "a", "K2" -> "k", "_13_" -> "c", "_15_" -> "e", "_18_" -> "l", "r" -> "", "_Signal" -> "s_", "_Detection p-value" -> "p_",
      "_Detection" -> "d_"}] &/@ columnNames;
dataPattern2 = ToExpression[dataPattern2]
```

```
{ac3s_, ac3d_, ac3p_, ae3s_, ae3d_, ae3p_, al3s_, al3d_, al3p_}
```

```
columnNames = StringDrop[#, StringPosition[#, "_"][[3, 1]]] &/@ Drop[Drop[data3[[1]], -1], 1];
dataPattern3 =
  StringReplace[#, {"T15" → "t", "C" -> "c", "_r" -> "", "_Signal" -> "s_", "_Detection p-value" -> "p_", "_Detection" -> "d_", "A311" -> "m", "L" -> "l",
      "K2" -> "k", "A297nb" -> "a", "T" -> "t", "E" -> "e", "1_2" -> "1"}] &/@ columnNames;
dataPattern3 = ToExpression[Append[dataPattern3, "description_"]]
```

```
{tc1s_, tc1d_, tc1p_, ml2s_, ml2d_, ml2p_, kt1s_, kt1d_, kt1p_, tt1s_, tt1d_, tt1p_, at1s_, at1d_, at1p_, mt1s_, mt1d_, mt1p_, tc3s_, tc3d_, tc3p_, te3s_, te3d_, te3p_,
 tl2s_, tl2d_, tl2p_, tl3s_, tl3d_, tl3p_, mc2s_, mc2d_, mc2p_, tc2s_, tc2d_, tc2p_, mc3s_, mc3d_, mc3p_, me3s_, me3d_, me3p_, ml3s_, ml3d_, ml3p_, kt2s_, kt2d_,
 kt2p_, kt3s_, kt3d_, kt3p_, tt2s_, tt2d_, tt2p_, tt3s_, tt3d_, tt3p_, at2s_, at2d_, at2p_, at3s_, at3d_, at3p_, mt2s_, mt2d_, mt2p_, te1s_, te1d_, te1p_, mt3s_,
 mt3d_, mt3p_, te2s_, te2d_, te2p_, tl1s_, tl1d_, tl1p_, mc1s_, mc1d_, mc1p_, me1s_, me1d_, me1p_, me2s_, me2d_, me2p_, ml1s_, ml1d_, ml1p_, description_}
```

```
TableForm[
  Sort[
   Transpose[
    {StringReplace[#, {"T15" → "t", "C" -> "c", "_r" -> "", "_Signal" -> "s_", "_Detection p-value" -> "p_", "_Detection" -> "d_", "A311" -> "m",
        "L" -> "l", "K2" -> "k", "A297nb" -> "a", "T" -> "t", "E" -> "e", "1_2" -> "1"}] &/@ columnNames, Drop[data3[[1]], -1]}]]];
```

```
dataPattern = Join[dataPattern1, dataPattern2, dataPattern3]
```

```
{identifier_, ae1s_, ae1d_, ae1p_, al1s_, al1d_, al1p_, kc1s_, kc1d_, kc1p_, ke1s_, ke1d_, ke1p_, kl1s_, kl1d_, kl1p_, ac2s_, ac2d_, ac2p_, ae2s_, ae2d_,
 ae2p_, al2s_, al2d_, al2p_, kc2s_, kc2d_, kc2p_, ke2s_, ke2d_, ke2p_, kl2s_, kl2d_, kl2p_, ac1s_, ac1d_, ac1p_, kc3s_, kc3d_, kc3p_, ke3s_, ke3d_,
 ke3p_, kl3s_, kl3d_, kl3p_, ac3s_, ac3d_, ac3p_, ae3s_, ae3d_, ae3p_, al3s_, al3d_, al3p_, tc1s_, tc1d_, tc1p_, ml2s_, ml2d_, ml2p_, kt1s_, kt1d_,
 kt1p_, tt1s_, tt1d_, tt1p_, at1s_, at1d_, at1p_, mt1s_, mt1d_, mt1p_, tc3s_, tc3d_, tc3p_, te3s_, te3d_, te3p_, tl2s_, tl2d_, tl2p_, tl3s_, tl3d_,
 tl3p_, mc2s_, mc2d_, mc2p_, tc2s_, tc2d_, tc2p_, mc3s_, mc3d_, mc3p_, me3s_, me3d_, me3p_, ml3s_, ml3d_, ml3p_, kt2s_, kt2d_, kt2p_, kt3s_, kt3d_,
 kt3p_, tt2s_, tt2d_, tt2p_, tt3s_, tt3d_, tt3p_, at2s_, at2d_, at2p_, at3s_, at3d_, at3p_, mt2s_, mt2d_, mt2p_, te1s_, te1d_, te1p_, mt3s_, mt3d_,
 mt3p_, te2s_, te2d_, te2p_, tl1s_, tl1d_, tl1p_, mc1s_, mc1d_, mc1p_, me1s_, me1d_, me1p_, me2s_, me2d_, me2p_, ml1s_, ml1d_, ml1p_, description_}
```

```
Length[dataPattern]
```

```
146
```

▪ Reject control initial lines

```
data = Drop[data, 57]; Dimensions[data]
```

```
{15866, 146}
```

▪ Reject absents (keep genes with at least 1 present value)

```
data = Cases[data, x_ /; Length[Cases[x, "P"]] > 0];
Dimensions[data]
```

```
{11859, 146}
```

```
Table[med[i] = Median[data[[All, i]]], {i, 2, Length[data[[1]]] - 2, 3}]
```

```
{351.5, 350.2, 411.4, 465.4, 269.6, 469., 495.9, 480.5, 456.8, 445.1, 439.8, 558.8, 447.8, 347.9, 497.5, 481.7, 336.1, 388.4, 327.1, 255.3, 279.6, 281.9, 260.5, 315.3, 288.3,
 255.5, 242.5, 293.3, 311.4, 210., 313.5, 378.1, 309.8, 318.8, 224.5, 318.4, 260.2, 173.8, 351.8, 281.4, 233.8, 335.1, 275.3, 265.5, 227.2, 176., 246.1, 285.3}
```

```
Take[data[[1]], 2]
```

```
{1398749_at, 7844.7}
```

```
Do[Do[data[[j, i]] = data[[j, i]] / med[i], {i, 2, Length[data[[1]]] - 2, 3}], {j, Length[data]}];
```

```
Take[data[[1]], 2]
```

```
{1398749_at, 22.3178}
```

```
fold1 = 2.5
```

```
2.5
```

```
fold2 = 2.5
```

```
2.5
```

```
pVal1 = 0.05
```

```
0.05
```

```
pVal2 = 0.05
```

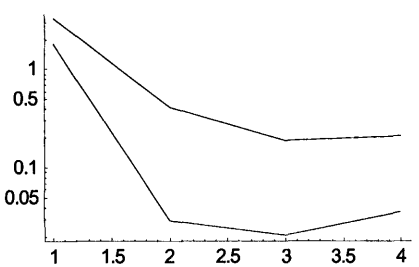
```
0.05
```

```
identifierSubList =
 Cases[fpData,
  {identifier_String, data__} :>
   identifier /;
    (((Cases[Take[Sort[Take[{data}, 6]], 5], {x_Real, y_} /; x > pVal1 || y < fold1] == {}) &&
        (Cases[Take[Sort[Take[{data}, 6]], {6, 6}], {x_Real, y_} /; x > pVal2 || y < fold1] == {}) && (Min[Cases[Take[{data}, 6], {x_, y_} /; x < pVal1 → y]] > 1)) ||
      ((Cases[Take[Sort[Take[{data}, 6]], 5], {x_Real, y_} /; x > pVal1 || y ≥ 1 / fold1] == {}) &&
        (Cases[Take[Sort[Take[{data}, 6]], {6, 6}], {x_Real, y_} /; x > pVal2 || y ≥ 1 / fold1] == {}) && (Max[Cases[Take[{data}, 6], {x_, y_} /; x < pVal1 → y]] < 1))) &&
    (Min[(Take[{data}, -4] /. {x_, y_} → x)] > pVal1)];
Length[identifierSubList]
```

```
cl = clusterAnalysis[clusterData];

dataSubSet = identifierSubList /. id_String:→ (Cases[fpData, {id, __}])[[1]];
Dimensions[dataSubSet]

colourOffset = -0.7; colourGain = 1.6;
TableForm[
 subSetTable =
  Prepend[
   Flatten[
    {First[#],
     (StyleForm[ToString[(#[[2]] * 10 // Round) / 10 // N] <> Which[#[[1]] < 0.01, "***", #[[1]] < 0.05, "**", #[[1]] < 0.1, "*", True, ""], FontWeight -> "Bold",
        FontFamily -> "Ariel", FontColor → hallNW[(#[[2]] + colourOffset) * colourGain]]) & /@ Rest[#], shortAnnot[First[#]]}] & /@ dataSubSet,
   {"identifier", "t/k", "a/k", "m/k", "tt/kt", "at/kt", "mt/kt", "kt/k", "tt/t", "at/a", "mt/m", "anottation"}], TableSpacing -> {1, 1}]
```

identifier	t/k	a/k	m/k	tt/kt	at/kt	mt/kt	kt/k	tt/t	at/a	mt/m	anottation
1377375_at	9.2**	16.8**	28.8***	33.8**	113.9***	59.**	0.1*	0.4**	0.7	0.2**	Similar to aminoadipate-semialdehyde synthase (LOC296925), mRNA
1368685_at	54.***	76.6***	64.9***	12.8***	14.6***	15.3***	6.2	1.5	1.2	1.5	Cspg4 membrane-spanning proteoglycan NG2
1370477_at	24.8***	49.6***	66.2***	3.9**	9.9**	6.8***	4.4**	0.7*	0.9	0.5**	Ocm oncomodulin
1371132_a_at	18.1***	17.3***	21.3***	13.6**	21.7**	24.1**	1.5	1.1	1.8	1.6	Ank3 ankyrin 3 (G)
1389353_at	0.1***	0.1**	0.1***	0.1**	0.1**	0.1***	15.8**	13.1*	11.5*	15.2*	
1368512_a_at	0.***	0.1***	0.***	0.**	0.***	0.**	11.9**	30.*	1.2	2.5**	Enpep aminopeptidase A
1373727_at	0.1***	0.1**	0.1***	0.1***	0.**	0.**	16.2*	19.9***	4.8	4.1*	BLASTx Hypothetical protein HEMBA1007053
1373748_at	0.***	0.***	0.***	0.1***	0.1***	0.1**	1.8*	13.9***	9.1	6.**	
1367581_a_at	7.2***	5.2***	5.8***	3.2***	3.**	2.6***	3.3**	1.5	1.9*	1.5*	Spp1 secreted phosphoprotein 1
1390474_at	2.9***	2.9***	3.2***	3.6**	2.8***	3.5***	1.3	1.4	1.1	1.3	Similar to RIKEN cDNA 9330161F08 (LOC362898), mRNA
1373642_at	5.7***	3.7***	6.1***	3.6**	4.3***	4.1**	1.2	0.8	1.4*	0.8	Similar to hypothetical protein (LOC310628), mRNA
1387153_at	4.1***	4.5***	4.9***	3.1**	3.5**	3.9**	1.4	1.1	1.1	1.1	Ril reversion induced LIM gene
1370418_s_at	3.6***	4.2***	5.9***	3.5***	4.7***	4.6***	1.4*	1.3	1.5*	1.1	Bk B/K protein
1368223_at	0.2**	0.3**	0.2**	0.3**	0.1***	0.1**	6.*	7.5	2.	4.7*	Adamts1 a disintegrin-like and metalloprotease (reprolysin type) with thrombo
1369044_a_at	0.2**	0.1**	0.1**	0.1**	0.1***	0.1**	10.2**	3.5	5.2	3.7	Pde4b phosphodiesterase 4B
1374157_at	0.1***	0.1***	0.1***	0.1***	0.1***	0.***	9.9*	5.3*	3.6**	3.3**	
1391428_at	0.1***	0.3**	0.4**	0.1**	0.1**	0.1**	10.2*	7.*	3.3	3.2	
1372101_at	0.4**	0.1**	0.2**	0.2***	0.1**	0.1**	9.1***	3.9	11.2	7.1*	Ppap2b ER transmembrane protein Dri 42
1372615_at	0.2**	0.2**	0.2**	0.2**	0.2**	0.2**	8.**	10.	9.6	6.	
1375267_at	0.***	0.2***	0.2***	0.2**	0.4**	0.2**	1.	10.6*	1.5**	1.2	Similar to Chain A, Cyclophilin C Complexed With Cyclosporin A (LOC291463), m
1374081_at	0.2***	0.***	0.1***	0.1**	0.**	0.***	2.9**	1.	3.5	2.2	
1374474_at	0.4**	0.1***	0.1***	0.2**	0.1***	0.1***	3.**	1.5	4.3	2.7	
1398823_at	0.3**	0.1**	0.1***	0.4***	0.4***	0.3**	0.8	1.1	3.7	2.	Tsnax translin-associated factor X
1398387_at	0.3**	0.1***	0.3**	0.2***	0.2***	0.2**	2.2	1.4	3.6**	1.1	MGC72614 Unknown (protein for MGC:72614)
1368565_at	0.3**	0.2**	0.3**	0.4**	0.4***	0.3**	1.3	1.4**	2.5**	1.3	Slc1a3 solute carrier family 1 (glial high affinity glutamate transporter), m

```
Show[
 {LogListPlot[
   (Cases[data, {identifierSubList[[8]], __}] /.
      dataPattern→ {Mean[Mean /@ {{kc1s, kc2s, kc3s}, {ke1s, ke2s, ke3s}, {kl1s, kl2s, kl3s}}], Mean[Mean /@ {{tc1s, tc2s, tc3s}, {te1s, te2s, te3s}, {tl1s, tl2s, tl3s}}],
        Mean[Mean /@ {{ac1s, ac2s, ac3s}, {ae1s, ae2s, ae3s}, {al1s, al2s, al3s}}], Mean[Mean /@ {{mc1s, mc2s, mc3s}, {me1s, me2s, me3s}, {ml1s, ml2s, ml3s}}]])[[1]],
  PlotJoined→ True, DisplayFunction→ Identity],
 LogListPlot[
  (Cases[data, {identifierSubList[[8]], __}] /.
     dataPattern→ {Mean[{kt1s, kt2s, kt3s}], Mean[{tt1s, tt2s, tt3s}], Mean[{at1s, at2s, at3s}], Mean[{mt1s, mt2s, mt3s}]})[[1]], PlotJoined→ True,
  DisplayFunction→ Identity]}, PlotRange→ All, DisplayFunction→ $DisplayFunction]
```



- Graphics -

Appendix 3: First microarray experiment: Genes differentially expressed between non-metastasising and metastasising cells.

*In this table, the Affymetrix Identifiers, expression ratios and symbols for each differentially expressed transcript are displayed. The Affymetrix Identifier is the unique identification for each transcript that is probed by the array. Not all transcripts are yet characterised, which is why only some of the Affymetrix Identifiers in this table have accompanying gene symbols. Full details of the genes in the following list can be accessed by entering the Affymetrix ID into the NetAffx™ Analysis Center at http://www.affymetrix.com/analysis/index.affx. Each column describes the expression ratio (ER) of two experimental groups, where ac = control A297 cells; ae = early (30 minute) response to PDGF/IGF of A297 cells; al = late response (3 hours) of A297 cells; kc = control K2 cells; ke = early (30 minute) response to PDGF/IGF of K2 cells; al = late response (3 hours) of K2 cells. High expression ratios (ER > 1) indicate that the numerator (the upper part of the fraction) is over-expressed in metastatic cells. Low expression ratios (ER < 1) indicate that the denominator (lower part of the fraction) is under-expressed in metastatic cells. Pseudocolours represent the degree of overexpression of the numerators, with red being high overexpression, through orange, yellow, green and blue with black representing underexpression of the numerator. Statistical significance is indicated as follows; *P < 0.1, **P < 0.05, ***P < 0.01.*

Affymetrix ID	ae/ac	al/ae	ke/kc	kl/ke	ac/kc	ae/ke	al/kl	
1377018 _at	1.	0.8	2.4	0.4	39.3	16.2	31.5	
1372013 _at	0.9	0.9	1.3	0.5	11.	7.7	14.5	
1372200 _at	1.	1.	0.6	0.9	8.5	12.6	14.8	
1377375 _at	1.	0.6	0.9	0.5	16.	17.4	19.7	
1371679 _at	0.7	0.8	1.	0.8	18.6	13.	12.9	
1372447 _at	0.9	0.9	1.1	1.	16.3	13.	12.4	Fgfr1
1368103 _at	1.2	0.9	1.2	1.2	16.1	16.8	13.5	Abcg1
1388866 _at	0.9	1.	0.8	1.3	18.1	19.4	15.	
1386935 _at	0.3	0.6	18.7	0.	2.9	0.1	0.8	Nr4a1
1369067 _at	0.2	1.8	24.8	0.1	5.7	0.	0.6	Nr4a3
1373776 _at	0.4	9.4	1.2	0.9	0.1	0.	0.3	
1389402 _at	2.5	0.5	5.2	0.2	1.	0.5	1.	
1387766 _a_at	0.4	4.6	1.	0.9	0.1	0.	0.2	Rbp2
1372539 _at	0.8	2.8	1.1	1.	0.	0.	0.	
1373748 _at	0.7	2.6	1.1	0.9	0.	0.	0.	
1368311 _at	0.4	3.3	1.	0.9	0.	0.	0.	Mgmt
1375729 _at	0.4	3.1	1.1	1.	0.1	0.	0.2	
1370503 _s_at	0.4	3.9	1.	1.1	0.	0.	0.	Epb4 .113
1368515 _at	1.	3.5	1.	1.	0.	0.	0.	Epb4 .113
1374318 _at	1.3	3.5	1.1	0.9	0.	0.	0.	
1374816 _at	0.9	0.9	1.	2.3	0.8	0.7	0.3	
1388924 _at	1.1	0.9	0.6	1.8	0.2	0.3	0.2	
1376667 _at	1.2	0.9	0.8	2.2	0.3	0.4	0.2	Cyp26b1
1373122 _at	1.	0.4	0.8	1.2	0.4	0.5	0.1	
1368106 _at	2.1	0.6	1.1	1.2	0.	0.	0.	Plk2
1371517 _at	1.8	0.8	0.9	1.	0.	0.	0.	
1370692 _at	1.9	0.5	0.9	1.1	0.1	0.3	0.1	Il1rl1
1372006 _at	1.7	0.5	0.9	1.2	0.2	0.4	0.2	
1387273 _at	1.5	0.7	1.	1.4	0.2	0.3	0.1	Il1rl1
1386899 _at	1.6	0.8	1.2	1.	0.1	0.2	0.1	
1370362 _at	1.3	0.6	1.	1.	0.1	0.1	0.1	Ptprn
1398823 _at	1.5	0.6	0.9	1.	0.1	0.1	0.1	Tsnax
1369484 _at	1.5	0.6	0.9	1.1	0.1	0.1	0.1	Wisp2
1376105 _at	1.5	1.1	1.	0.5	0.	0.1	0.1	
1368223 _at	0.7	1.2	1.5	0.8	0.4	0.2	0.3	Adamts1
1376937 _at	0.8	0.8	1.4	0.5	0.4	0.3	0.4	
1387951 _at	0.6	0.8	1.5	0.8	0.4	0.2	0.2	Daf
1372750 _at	0.7	0.6	1.7	0.8	0.3	0.1	0.1	
1367888 _at	1.	1.1	1.1	1.4	0.1	0.1	0.1	Pcdh21
1367759 _at	1.1	0.7	1.	0.8	0.3	0.3	0.3	H1f0
1374728 _at	1.1	0.6	1.	0.6	0.2	0.2	0.2	
1374683 _at	0.8	0.8	1.1	0.8	0.3	0.2	0.2	
1374953 _at	0.9	0.8	1.3	0.8	0.2	0.2	0.1	
1374029 _at	0.6	1.1	0.9	0.9	0.4	0.2	0.3	
1376158 _at	1.	0.7	1.1	1.1	0.2	0.2	0.1	
1368541 _at	1.	1.	0.8	1.	0.1	0.2	0.2	Emb
1387136 _at	1.1	1.	1.	0.8	0.2	0.2	0.3	Esp
1369948 _at	1.	1.	1.1	0.8	0.2	0.2	0.2	Ngfrap1
1376786 _a_at	0.9	1.1	0.9	1.	0.2	0.2	0.2	
1372323 _at	0.9	0.9	1.	0.9	0.3	0.2	0.2	Sardh
1375267 _at	1.	1.	1.	1.	0.2	0.2	0.2	
1368021 _at	0.9	1.	0.9	1.	0.2	0.2	0.2	Adh1
1370609 _a_at	0.9	0.9	1.	0.9	0.2	0.2	0.2	Slc16a7
1386938 _at	0.8	1.	1.	1.	0.1	0.1	0.1	Anpep
1370959 _at	0.7	0.9	1.	1.	0.2	0.1	0.1	Col3a1
1374705 _at	0.7	0.8	0.9	1.	0.1	0.1	0.1	
1370161 _at	0.9	0.9	1.1	0.9	0.1	0.1	0.1	Ssg1
1371475 _at	1.	0.9	1.	1.1	0.1	0.1	0.1	Rnase4
1370026 _at	1.	0.9	1.	1.	0.	0.	0.	
1370301 _at	1.1	0.9	1.	1.	0.	0.	0.	Mmp2

1369652_at	1.2	1.1	1.	1.1	0.2	0.2	0.2	
1373490_at	1.3	0.9	1.	1.	0.2	0.2	0.2	
1389503_at	1.2	0.8	1.	0.9	0.2	0.2	0.2	
1368990_at	1.3	1.	0.9	0.9	0.1	0.1	0.1	Cyp1b1
1368281_at	1.2	1.	1.1	1.	0.1	0.1	0.1	Dpep1
1368304_at	1.2	1.	1.	0.9	0.1	0.1	0.1	Fmo3
1388583_at	1.2	0.9	1.	1.1	0.1	0.1	0.1	
1390638_at	1.2	0.9	0.9	1.1	0.	0.	0.	
1368674_at	1.1	1.5	1.	0.9	0.1	0.1	0.2	Pygl
1376047_at	0.9	1.6	1.	0.9	0.1	0.1	0.1	
1369651_at	0.8	1.4	1.	1.	0.2	0.2	0.2	
1387709_at	0.8	1.4	1.	0.9	0.1	0.1	0.2	Figf
1375053_at	0.9	1.4	1.	0.9	0.1	0.1	0.1	
1368259_at	0.9	1.2	1.	1.1	0.1	0.1	0.1	Ptgs1
1374622_at	0.8	1.2	1.	1.	0.1	0.1	0.1	
1371688_at	1.	1.4	0.9	1.1	0.2	0.2	0.2	
1390397_at	0.9	1.2	0.9	1.1	0.3	0.2	0.3	
1390468_at	0.8	1.2	0.9	1.1	0.3	0.2	0.2	
1387669_a_at	0.8	1.3	0.7	0.9	0.2	0.2	0.3	Ephx1
1369647_at	1.	1.2	0.8	1.	0.1	0.1	0.2	Calcrl
1370229_at	1.	1.2	1.	0.9	0.2	0.2	0.2	Ndr4
1387197_at	0.9	1.4	0.9	0.9	0.2	0.2	0.3	Omd
1387995_a_at	0.9	1.3	1.	1.	0.3	0.3	0.3	Ifitm2
1374081_at	0.4	1.2	0.9	1.	0.	0.	0.	
1381504_at	0.5	1.3	1.	1.	0.	0.	0.	
1369294_at	0.7	1.4	1.	0.9	0.	0.	0.	Bst1
1388821_at	0.9	1.2	1.1	0.7	0.1	0.1	0.1	
1371056_at	0.6	1.2	1.1	0.9	0.	0.	0.	Neo1
1371889_at	0.6	1.1	1.1	0.8	0.2	0.1	0.1	Slc22a17
1371765_at	1.3	0.7	0.9	1.1	0.4	0.5	0.3	
1372728_at	1.3	0.9	0.9	1.1	0.4	0.6	0.5	
1389681_at	1.3	1.3	0.9	1.1	0.3	0.5	0.6	
1375216_at	1.4	1.1	1.	1.	0.2	0.3	0.4	
1373198_at	1.	1.1	1.3	0.9	0.3	0.2	0.3	
1370642_s_at	0.7	1.	1.1	1.2	0.6	0.4	0.4	Pdgfrb
1373853_at	0.9	0.9	1.	0.7	0.6	0.5	0.7	
1369982_at	0.9	1.2	1.	0.9	0.6	0.5	0.7	Ap2a2
1398383_at	0.8	1.1	0.9	0.8	0.5	0.5	0.7	
1389349_s_at	1.1	0.9	0.9	0.8	0.3	0.4	0.4	Il17re
1370813_at	1.1	0.9	1.1	0.9	0.3	0.3	0.3	Gstm5
1390730_at	1.2	0.9	1.2	0.8	0.3	0.3	0.4	
1374656_at	1.	0.8	1.2	0.9	0.6	0.5	0.5	Sec6
1382672_a_at	0.9	1.1	1.1	0.8	0.5	0.4	0.5	
1370805_at	0.8	1.3	1.	0.9	0.4	0.3	0.4	Msg1
1375862_at	0.8	1.1	1.	0.9	0.4	0.3	0.4	
1371419_at	0.9	1.2	0.9	0.9	0.4	0.4	0.5	Spnb2
1388893_at	0.9	1.2	1.	0.9	0.3	0.3	0.4	
1398943_at	1.1	1.	1.	0.9	0.4	0.4	0.5	
1367864_at	1.1	1.	1.1	0.9	0.3	0.3	0.4	Pfkm
1389229_at	1.1	0.8	1.	0.9	0.4	0.4	0.4	
1370287_a_at	0.9	0.9	1.1	1.1	0.5	0.4	0.3	Tpm1
1376319_at	1.	1.	1.	1.1	0.4	0.4	0.4	
1373463_at	0.9	0.9	1.	0.9	0.3	0.3	0.3	
1367593_at	0.9	0.9	1.	0.9	0.4	0.4	0.4	Sepw1
1388628_at	1.	1.	1.	0.9	0.4	0.4	0.4	
1387854_at	1.1	1.	1.	1.1	0.5	0.5	0.5	Col1a2
1398349_at	1.2	1.	1.	1.1	0.5	0.5	0.5	Ak2
1388726_a_at	1.	1.	1.	1.	0.5	0.5	0.6	
1374290_at	1.	1.	1.	0.9	0.5	0.5	0.6	
1367987_at	0.9	1.	1.	1.	0.6	0.5	0.6	Rnpep

1371483_at	0.9	1.	1.	1.	0.6	0.6	0.6	Nnt
1389395_at	1.	1.	1.1	1.1	0.5	0.5	0.5	
1376339_at	1.	1.1	1.	1.	0.4	0.4	0.5	
1367744_at	1.	1.2	1.	0.9	0.5	0.4	0.5	Maged2
1376486_at	0.9	1.1	1.	1.	0.5	0.4	0.5	
1376736_at	0.7	1.8	0.8	1.2	0.	0.	0.	
1398348_at	0.7	2.1	1.1	0.9	0.1	0.	0.1	
1373577_at	0.5	1.9	1.1	0.9	0.	0.	0.1	
1371824_at	0.5	1.8	1.3	0.9	0.1	0.	0.	
1375026_at	0.4	1.6	1.3	0.8	0.1	0.	0.1	
1372715_at	0.7	1.6	1.1	1.	0.	0.	0.	
1386160_at	0.7	1.5	1.3	0.9	0.1	0.1	0.1	
1368146_at	0.4	1.2	0.8	1.	0.7	0.3	0.4	
1376362_at	0.9	1.4	1.1	1.3	0.4	0.3	0.3	
1376572_a_at	0.6	1.5	1.2	0.5	0.3	0.2	0.5	
1368358_a_at	0.6	1.4	1.	1.	0.5	0.3	0.4	Ptprr
1370971_at	0.7	1.5	1.1	0.8	0.3	0.2	0.4	MYHC
1373287_at	0.6	1.5	0.9	0.9	0.2	0.1	0.2	
1387018_at	0.8	1.8	1.3	0.8	0.4	0.3	0.6	Argbp2
1374193_at	0.7	1.8	1.1	1.1	0.5	0.3	0.5	
1387219_at	1.4	2.3	1.	0.8	0.2	0.2	0.7	Adm
1371988_at	1.5	1.5	0.9	1.2	0.1	0.1	0.2	
1377383_at	1.7	1.7	1.1	0.9	0.1	0.1	0.3	
1374855_at	1.	0.9	1.7	0.4	0.9	0.5	1.1	
1387870_at	1.	0.3	1.5	0.6	1.	0.6	0.4	Zfp36
1367802_at	0.9	0.6	1.6	0.6	1.2	0.7	0.7	Sgk
1372964_at	0.7	0.6	1.4	0.6	1.2	0.6	0.7	
1368489_at	1.6	1.	1.4	1.1	0.9	1.1	1.	Fosl1
1375061_at	1.2	0.8	1.5	0.8	0.7	0.6	0.6	
1373703_at	1.1	0.6	1.1	0.8	0.8	0.8	0.5	
1375382_at	1.1	0.5	1.1	0.7	0.9	0.9	0.7	Mdc1
1371680_at	0.8	0.7	1.	0.8	1.	0.8	0.7	
1374118_at	1.1	0.7	1.	1.1	0.8	0.8	0.6	
1388533_at	1.	0.8	1.	0.9	0.9	0.9	0.8	
1372769_at	1.	0.9	1.	1.	0.9	0.9	0.8	
1388690_at	1.	0.7	0.9	0.8	0.9	0.9	0.8	
1389720_at	0.9	0.7	0.9	0.9	0.8	0.8	0.6	
1372042_at	1.	1.3	0.9	1.1	0.6	0.6	0.8	
1391437_at	1.	1.3	0.9	1.1	0.6	0.7	0.8	
1371437_at	1.1	1.1	0.9	1.1	0.6	0.7	0.7	
1389973_a_at	0.9	1.2	1.	1.2	0.6	0.6	0.6	
1372169_at	1.	1.2	1.	1.2	0.8	0.8	0.8	
1373214_at	0.9	1.	0.8	1.1	0.7	0.8	0.8	
1374025_at	1.	1.1	0.9	1.1	0.7	0.8	0.8	
1371579_at	1.	1.	1.	0.9	0.8	0.7	0.8	
1398245_at	1.	0.9	1.	1.	0.6	0.7	0.6	
1371843_at	1.1	0.9	1.	1.	0.6	0.7	0.6	
1373933_at	1.	0.8	1.	1.	0.7	0.7	0.6	
1371774_at	1.1	0.8	1.1	1.	0.7	0.7	0.6	Sat
1389658_at	1.1	0.8	1.	1.	0.6	0.7	0.5	
1368835_at	1.	1.1	1.	0.8	0.6	0.6	0.8	Stat1
1370859_at	1.	1.	1.	0.9	0.6	0.6	0.7	P5
1389857_at	1.1	1.	0.9	1.	0.7	0.8	0.8	
1373870_at	1.2	1.1	1.	1.	0.6	0.7	0.7	
1371509_at	1.1	1.	1.	0.9	0.6	0.7	0.7	
1373185_at	1.1	1.	0.9	1.	0.6	0.7	0.7	
1387371_at	1.4	0.8	1.1	1.1	0.8	0.9	0.7	Cdc25a
1374331_at	1.2	1.	1.	1.2	0.9	1.	0.8	
1371424_at	0.9	1.3	0.8	1.5	1.	1.1	1.	
1375282_at	1.	1.2	1.3	1.2	1.1	0.9	0.9	

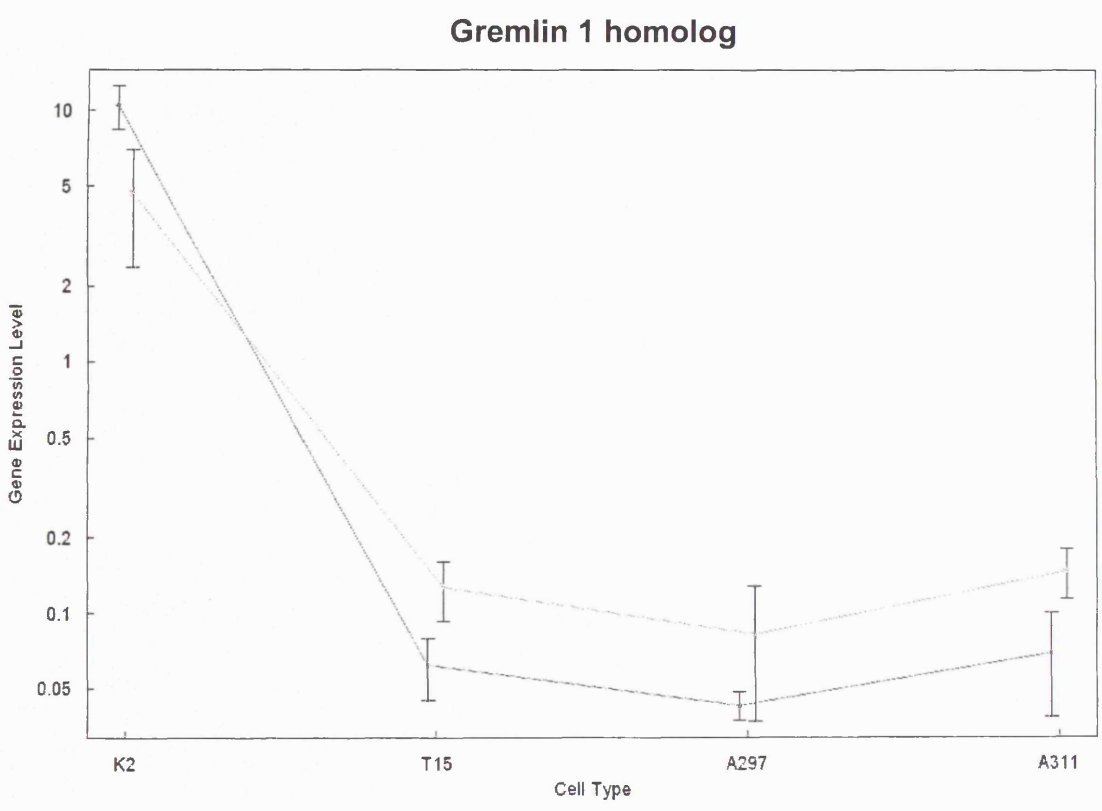
1374489 _at	0.8	1.2	1.1	0.7	0.9	0.7	1.1	
1398596 _at	1.	1.	1.2	0.8	1.1	0.8	1.1	
1375182 _at	0.7	1.5	0.9	1.	1.3	0.9	1.5	
1374719 _at	0.9	1.5	1.2	1.1	1.	0.8	1.1	
1369690 _at	0.8	1.3	1.	1.	1.2	1.	1.3	Nsf
1390222 _at	0.9	1.4	1.	1.1	1.1	1.	1.3	
1373751 _at	1.	1.9	1.1	0.9	0.6	0.5	1.1	
1388253 _at	0.9	1.7	0.9	1.1	0.8	0.7	1.1	
1389239 _at	0.8	1.6	0.7	1.1	0.8	0.8	1.2	
1367723 _a_at	0.9	1.5	1.	1.2	0.7	0.7	0.9	Lnk
1373206 _at	0.9	1.5	1.	1.1	0.7	0.6	0.8	
1392920 _at	0.8	1.4	1.	1.	0.7	0.6	0.8	
1368321 _at	2.3	0.1	2.6	0.4	0.6	0.5	0.1	Egr1
1372510 _at	1.7	0.8	1.9	0.4	0.6	0.6	1.1	
1388686 _at	1.3	1.	3.3	0.3	1.3	0.5	1.6	Dscr1
1373093 _at	1.7	0.9	3.1	0.5	1.2	0.7	1.3	
1368050 _at	1.5	0.5	2.4	0.4	1.6	1.	1.2	Ccnl1
1387788 _at	1.5	0.3	2.3	0.4	2.1	1.4	0.9	Junb
1372389 _at	1.4	0.4	2.8	0.4	2.	1.	0.9	
1398270 _at	1.1	1.4	1.6	0.6	2.8	1.9	4.3	Bmp2
1373719 _at	0.9	1.5	0.9	0.9	2.2	2.2	3.6	
1386976 _at	1.	1.2	1.	1.2	2.9	2.8	2.7	Kail
1372758 _at	0.9	1.3	1.	1.1	3.	2.7	3.1	
1388825 _at	0.9	1.2	1.2	1.1	2.9	2.3	2.4	
1390585 _at	0.9	1.	1.	0.8	2.9	2.5	3.	Maspl
1372069 _at	0.8	0.9	1.2	0.7	3.	2.2	2.8	
1390430 _at	0.9	0.8	1.2	0.7	3.	2.2	2.6	
1390429 _at	1.2	1.3	0.8	1.4	2.1	3.1	2.7	
1374403 _at	1.1	1.	0.8	1.	2.3	3.1	3.2	Efnb1
1390474 _at	1.	0.8	0.9	0.9	2.7	3.1	2.7	
1387808 _at	1.1	0.8	1.	0.9	2.4	2.8	2.7	Slc7a7
1376082 _at	1.2	0.7	1.	0.8	2.5	3.	2.8	
1371016 _at	1.3	0.6	0.9	0.8	1.7	2.7	1.9	
1374440 _at	0.8	1.	0.9	1.	2.6	2.3	2.2	
1389651 _at	1.1	1.	1.1	1.2	2.4	2.4	2.	
1373431 _at	1.	0.8	0.9	1.	2.2	2.3	2.	
1379371 _at	1.	0.8	1.	1.	2.3	2.3	1.9	
1376801 _at	1.2	1.	1.	1.1	2.2	2.6	2.5	
1373247 _at	0.9	0.9	0.8	1.1	2.4	2.7	2.1	
1389418 _at	1.	0.9	0.9	1.2	2.7	2.9	2.1	
1398381 _at	1.3	0.9	0.9	1.4	2.3	3.2	2.	
1376909 _at	1.2	0.7	0.9	1.3	2.3	2.9	1.6	
1386934 _at	1.8	0.7	1.	1.	1.4	2.4	1.6	
1371908 _at	1.6	0.7	0.9	1.	1.3	2.2	1.5	
1374864 _at	1.1	0.9	2.1	0.5	1.6	0.8	1.5	
1388842 _at	1.	1.2	1.5	0.7	1.8	1.2	2.2	
1388587 _at	1.3	0.8	1.7	0.8	2.2	1.7	1.8	Ler3
1375014 _at	0.9	1.1	1.7	0.9	2.4	1.3	1.8	
1373787 _at	1.	0.5	1.1	1.2	1.8	1.6	0.7	
1372665 _at	1.1	0.6	1.1	1.	1.3	1.3	0.7	
1372917 _at	1.3	0.5	1.2	0.9	1.2	1.3	0.8	
1387927 _a_at	0.7	1.	1.1	1.1	1.7	1.1	1.	Olfml
1390610 _at	0.9	1.	1.5	1.2	1.8	1.1	0.9	
1389549 _at	1.	1.2	0.9	1.	1.2	1.3	1.5	
1371087 _a_at	1.1	1.4	0.9	1.4	1.2	1.4	1.4	Mtap6
1368025 _at	1.1	1.3	0.8	1.3	1.2	1.6	1.6	Ddit4
1388505 _at	1.	1.4	0.8	1.3	1.3	1.5	1.6	
1371513 _at	1.	1.	1.	0.9	1.5	1.5	1.6	
1374987 _at	1.	1.2	1.	1.1	1.4	1.5	1.6	
1390463 _at	1.	1.2	1.	1.	1.6	1.6	1.8	

1370251_at	0.9	1.2	1.1	1.	1.4	1.2	1.4	Avpi1
1388978_at	0.9	1.1	0.9	1.2	1.4	1.6	1.5	
1367826_at	1.	1.1	1.	1.1	1.3	1.4	1.4	Nfe2l2
1371610_at	0.9	1.1	1.	1.1	1.5	1.4	1.4	
1389136_at	1.1	0.9	1.1	0.9	1.3	1.3	1.4	
1398294_at	1.1	0.9	1.1	1.	1.5	1.5	1.3	Actn1
1388474_at	1.	0.9	1.1	1.	1.5	1.4	1.2	
1372674_at	1.1	0.9	1.2	1.	1.5	1.4	1.3	
1367893_a_at	1.	0.8	1.	0.9	1.5	1.6	1.5	LOC170927
1386902_at	1.	1.	0.9	1.	1.3	1.4	1.4	Vdac3
1367485_at	1.	0.9	0.9	1.	1.4	1.5	1.4	
1372839_at	1.1	0.9	0.9	1.	1.4	1.6	1.4	
1390194_at	0.9	0.9	1.1	1.	1.3	1.2	1.1	
1389141_at	1.1	0.9	0.9	1.	1.1	1.3	1.2	
1371546_at	1.2	1.1	1.	1.2	1.1	1.3	1.2	
1383660_at	1.2	1.	1.1	1.1	1.1	1.3	1.1	
1368878_at	1.4	1.	1.4	0.7	1.2	1.2	1.7	Idi1
1388868_at	1.	1.	1.3	0.7	1.6	1.2	1.6	
1374446_at	1.	0.8	1.5	0.7	1.5	1.1	1.2	
1370956_at	0.6	1.7	0.9	1.3	2.4	1.7	2.4	Dcn
1367839_at	1.2	1.1	0.9	0.8	1.4	1.8	2.5	Fdft1
1389797_at	1.1	1.5	1.1	1.	2.1	2.	2.8	
1375965_at	0.9	1.2	0.9	1.1	2.	2.1	2.3	
1386969_at	1.1	1.3	0.9	1.1	1.8	2.	2.3	Nrn1
1371485_at	1.	1.3	0.9	1.1	1.8	2.	2.3	
1389017_at	1.	1.	0.9	0.9	1.9	2.1	2.3	
1373513_at	1.2	1.2	1.1	1.	1.9	2.1	2.5	
1379739_at	0.9	1.5	1.	1.1	1.9	1.6	2.1	
1372250_at	0.8	1.4	1.	1.2	2.3	1.9	2.1	
1374268_at	1.	1.4	1.2	1.1	2.3	1.9	2.2	
1370848_at	1.3	1.1	1.4	1.1	1.9	1.7	1.8	Slc2a1
1372347_at	0.9	1.2	1.1	0.9	1.9	1.4	1.8	
1372023_at	0.9	0.9	1.1	0.8	2.1	1.7	2.	
1375602_at	0.8	1.1	1.	0.7	2.	1.6	2.3	
1374228_at	0.9	1.1	1.	1.4	1.9	1.8	1.4	
1390684_at	1.1	0.9	1.1	1.1	1.9	1.9	1.6	
1398297_at	1.	1.3	0.9	1.2	1.8	1.8	2.	Mapk12
1384989_at	1.	1.2	0.8	1.4	1.7	2.	1.8	
1389103_at	1.1	0.9	1.	0.9	1.6	1.8	1.8	
1389551_at	1.1	1.	0.9	1.1	1.7	2.2	1.9	
1386908_at	1.	1.	0.9	1.1	1.8	1.9	1.7	Glrx1
1391981_at	1.	1.	0.9	1.1	1.8	2.	1.8	
1375120_at	0.6	0.9	1.	0.8	8.3	5.3	5.8	Idb4
1377092_at	1.3	1.	1.6	0.6	3.6	2.8	4.2	
1369338_at	0.5	1.6	1.	1.	4.	2.1	3.5	Robo1
1376943_at	0.8	0.8	1.2	0.7	4.2	2.7	3.	
1398305_at	0.9	0.9	0.7	1.7	3.2	3.9	2.2	Bk
1388632_at	0.9	1.	0.9	1.	3.6	3.6	3.7	
1398840_at	1.2	0.8	1.	1.2	3.8	4.6	3.3	Vamp5
1370023_at	0.9	1.	0.9	1.3	4.2	4.4	3.2	Gja4
1368657_at	1.1	1.	1.1	1.1	5.	5.1	4.4	Mmp3
1373494_at	1.	1.1	1.1	1.	4.5	3.9	4.4	
1370942_at	0.9	1.4	1.2	0.9	3.9	3.1	5.	
1371786_at	1.	1.	1.	0.8	3.8	4.	5.1	
1390221_at	0.8	0.9	0.7	2.2	4.8	5.7	2.2	
1376775_at	0.7	0.9	1.	1.5	6.3	4.7	2.9	
1368180_s_at	0.9	1.1	0.5	1.6	4.5	7.9	5.3	Gsta2
1374908_at	0.9	1.3	0.7	1.1	4.4	6.1	7.3	

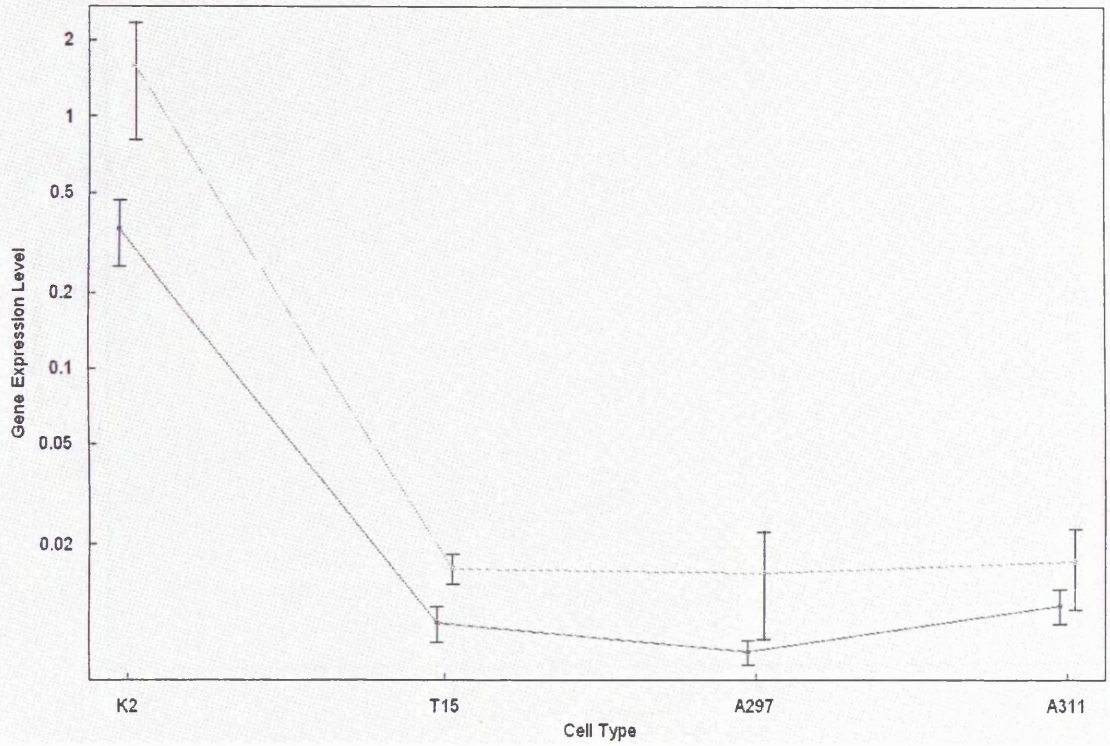
Appendix 4:

Expression patterns of candidate genes from the extended microarray experiment.

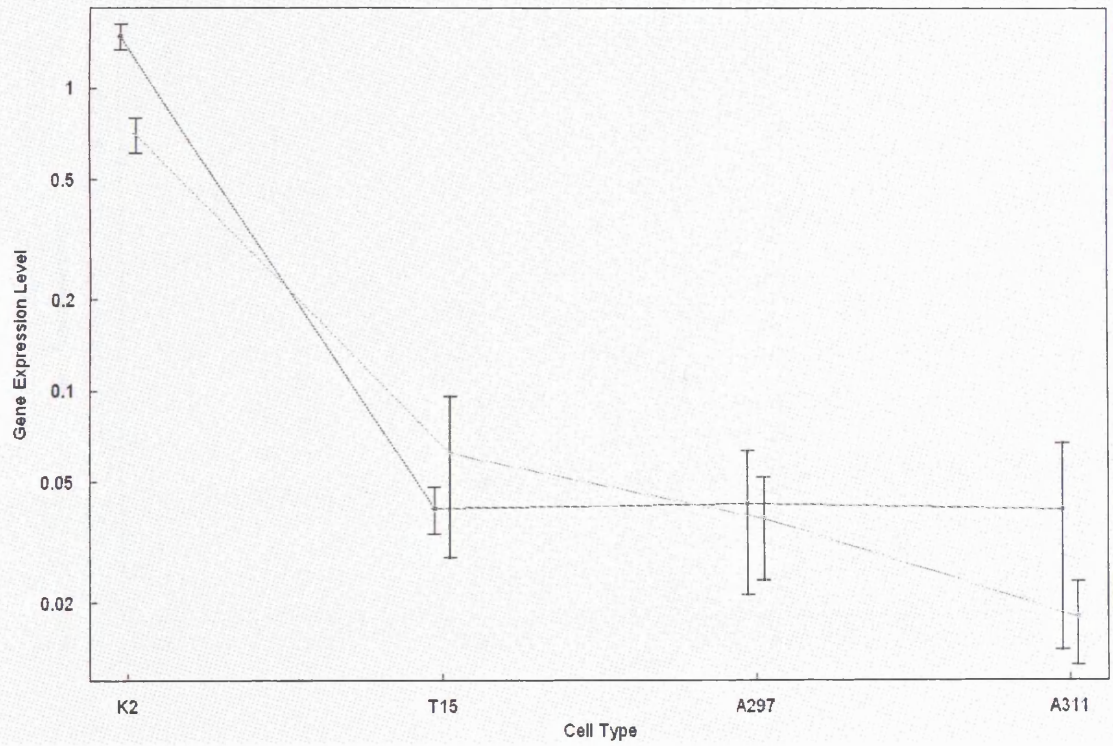
Absolute gene expression values are shown for the cultured cells and primary tumours, where the green line represents expression levels in the cultured cells, and the red line represents expression levels in the primary tumours. In the following graphs, the expression patterns of the cells and tumours are closely matched since this was one of the criteria used to obtain the list of candidate genes.

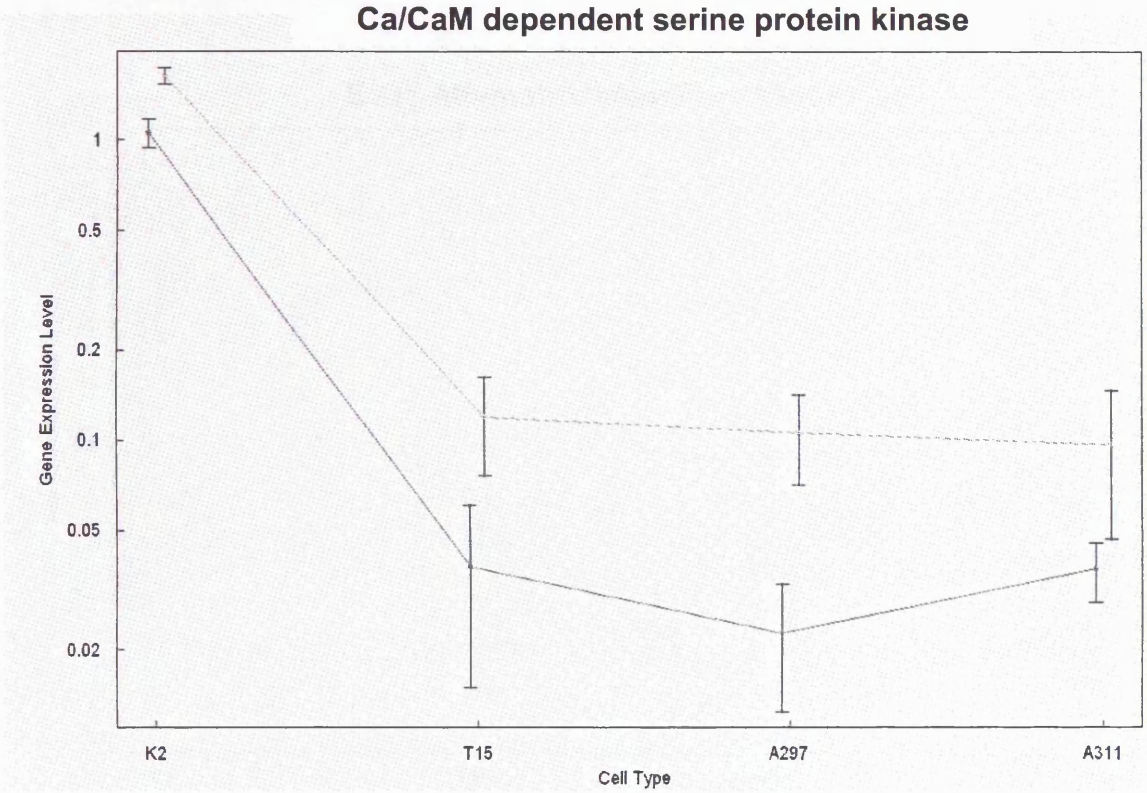


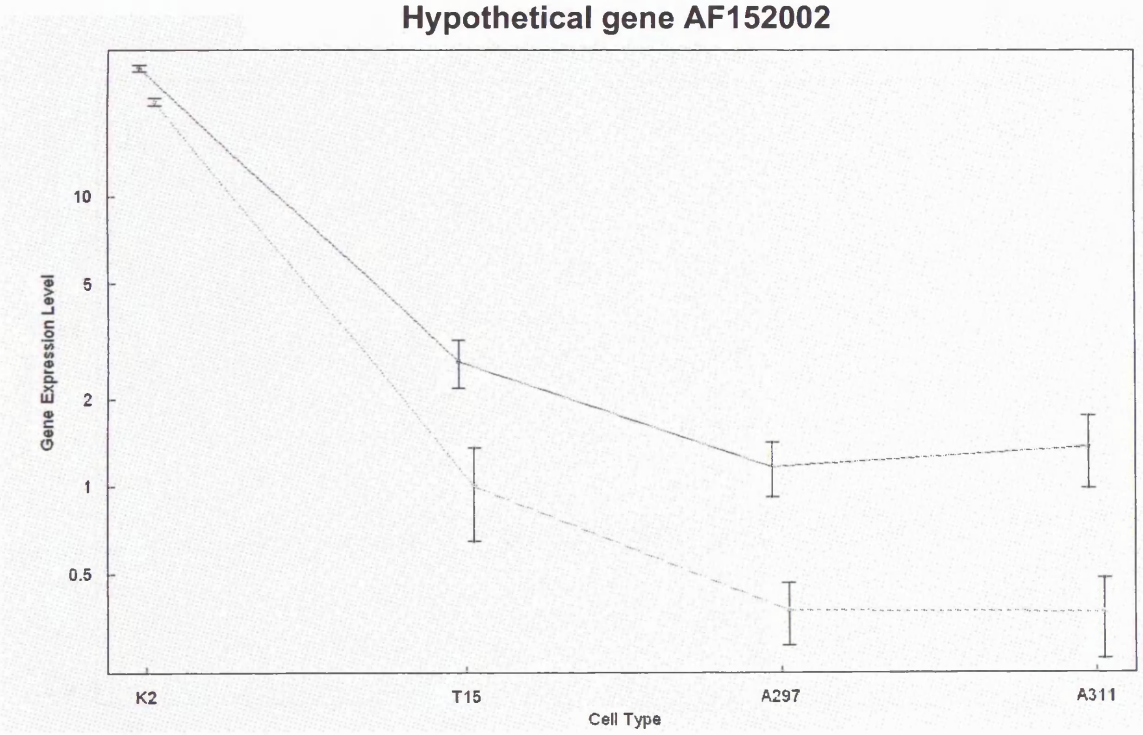
PAR-related orphan receptor α

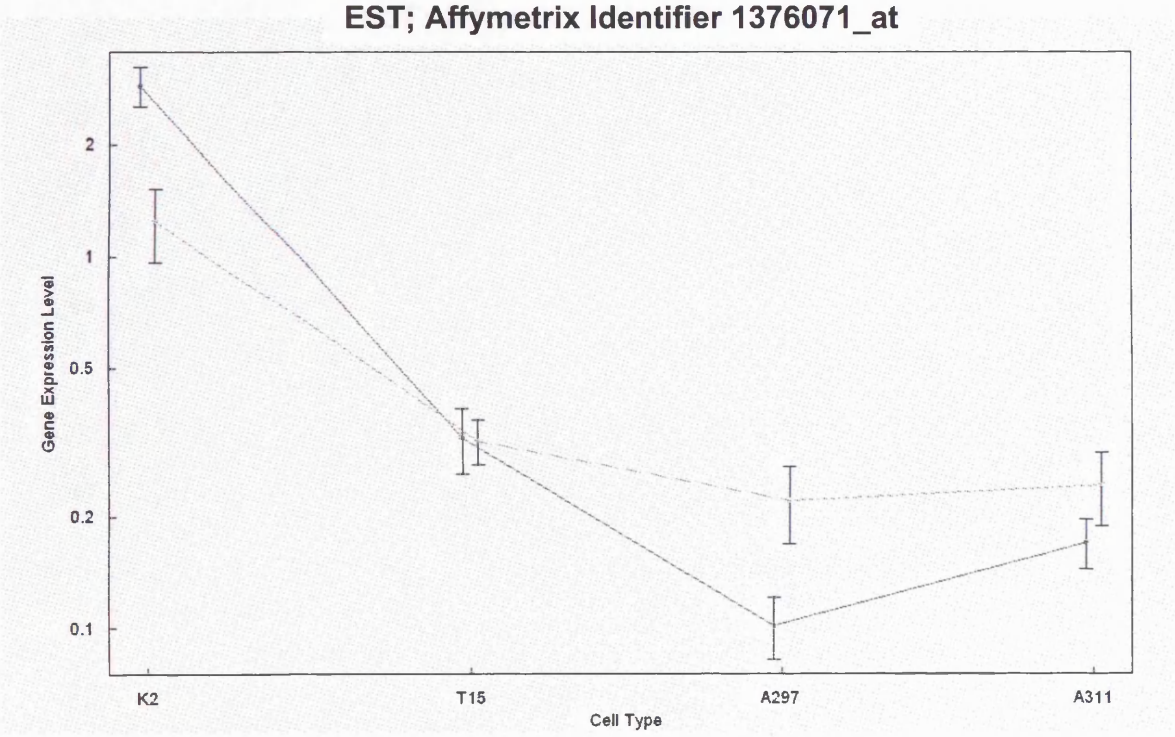


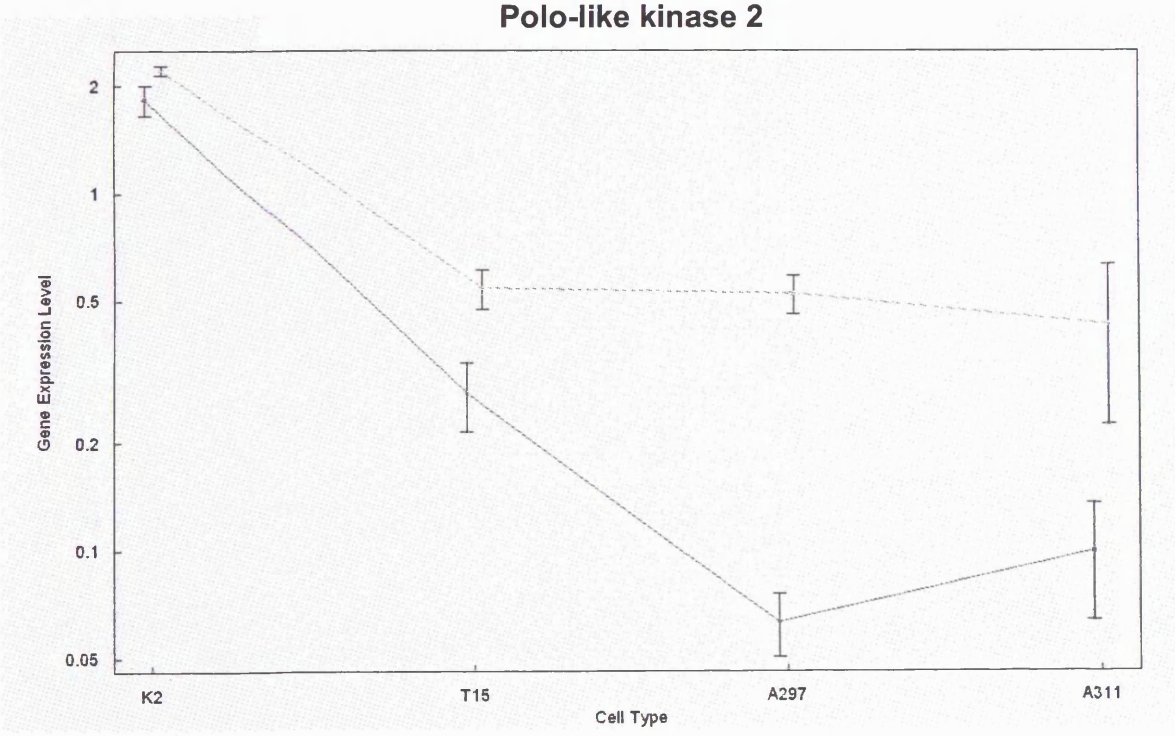
Erythrocyte protein band 4.1-like 3

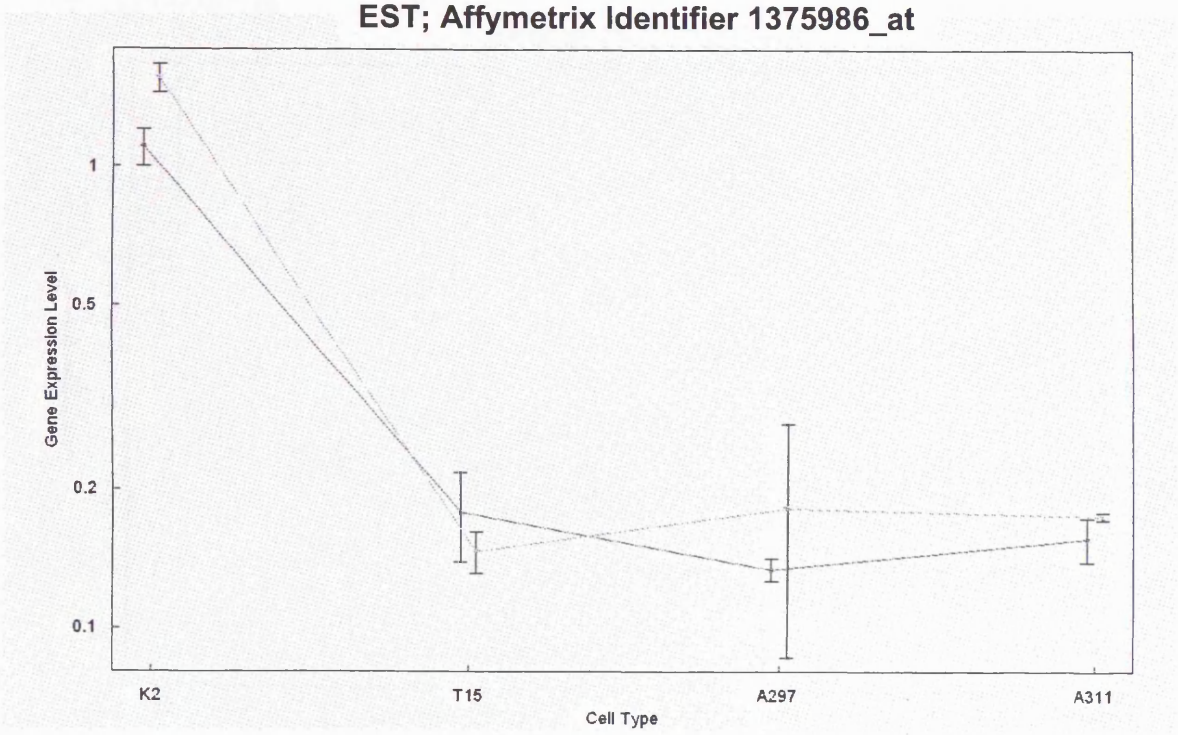


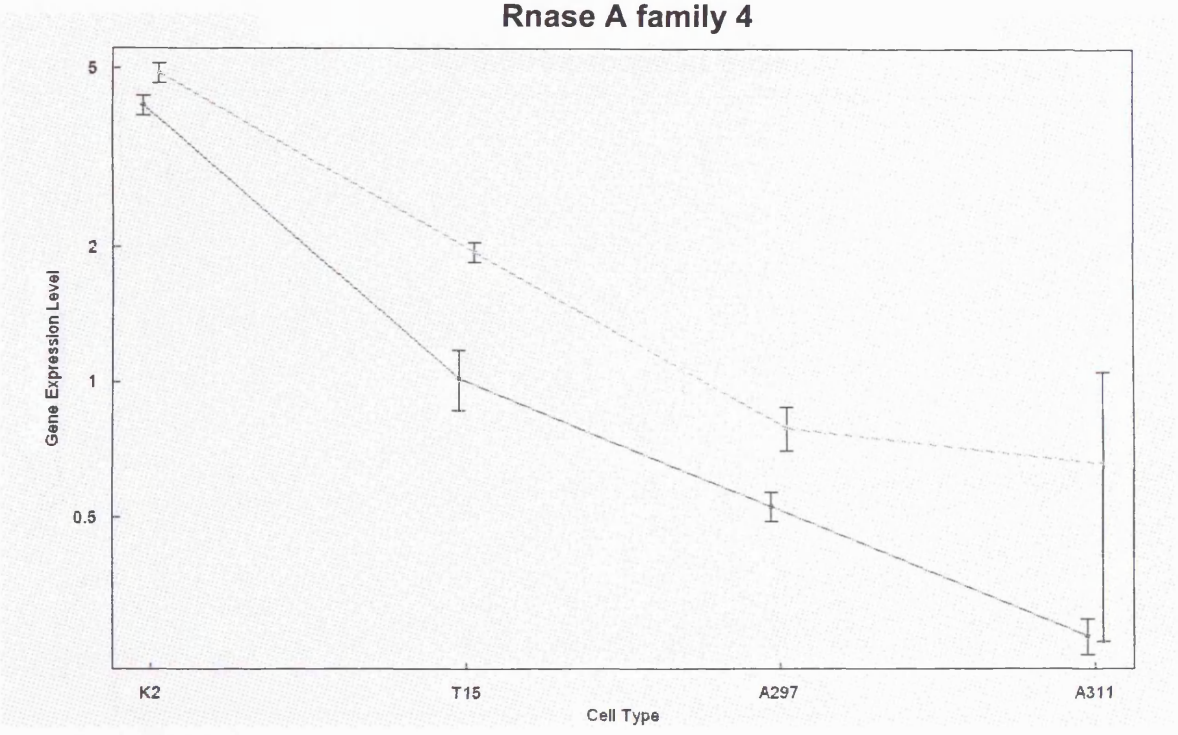

Ca/CaM dependent serine protein kinase
Gene Expression Level
1
0.5
0.2
0.1
0.05
0.02
K2
T15
A297
A311
Cell Type

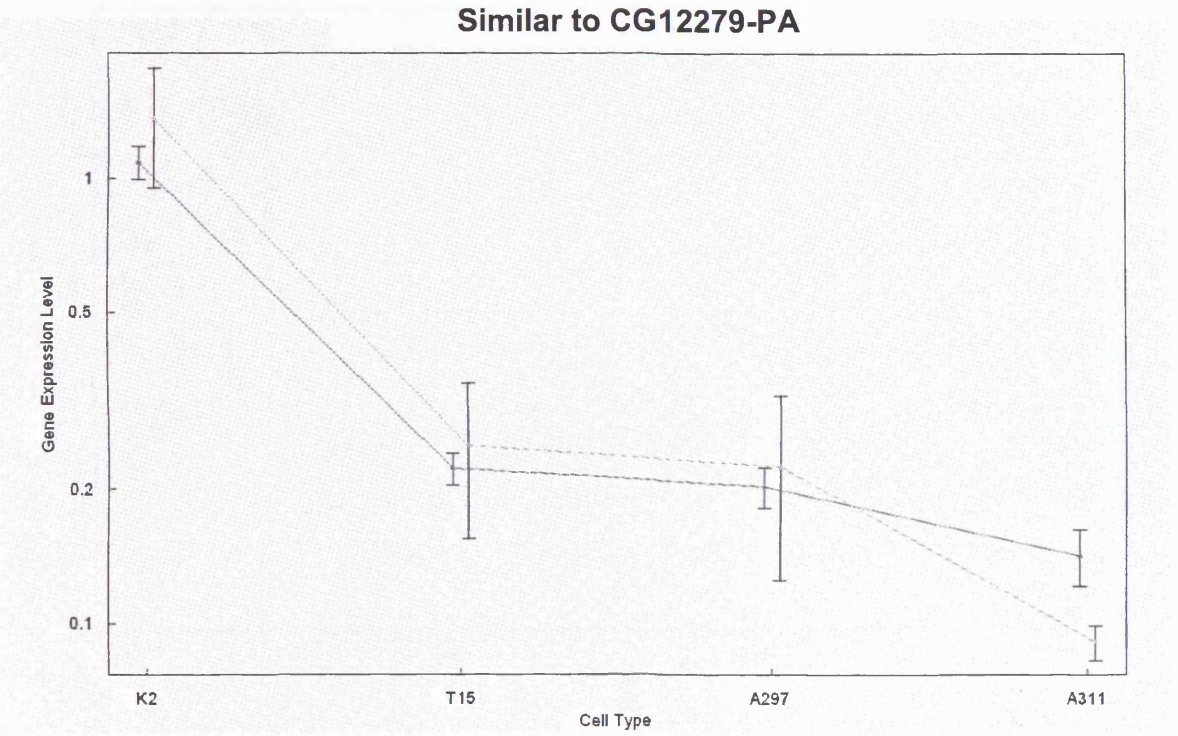

Hypothetical gene AF152002
Gene Expression Level
10
5
2
1
0.5
K2
T15
A297
A311
Cell Type

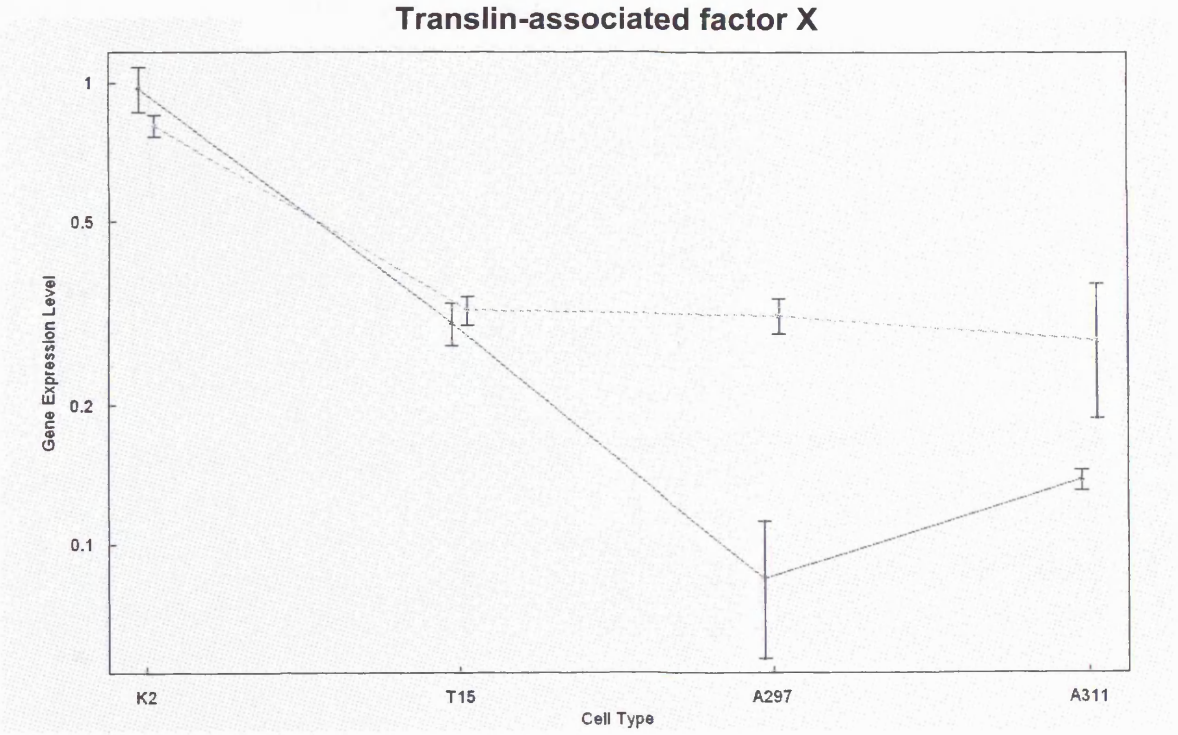

EST; Affymetrix Identifier 1376071_at
Gene Expression Level
2
1
0.5
0.2
0.1
K2
T15
A297
A311
Cell Type

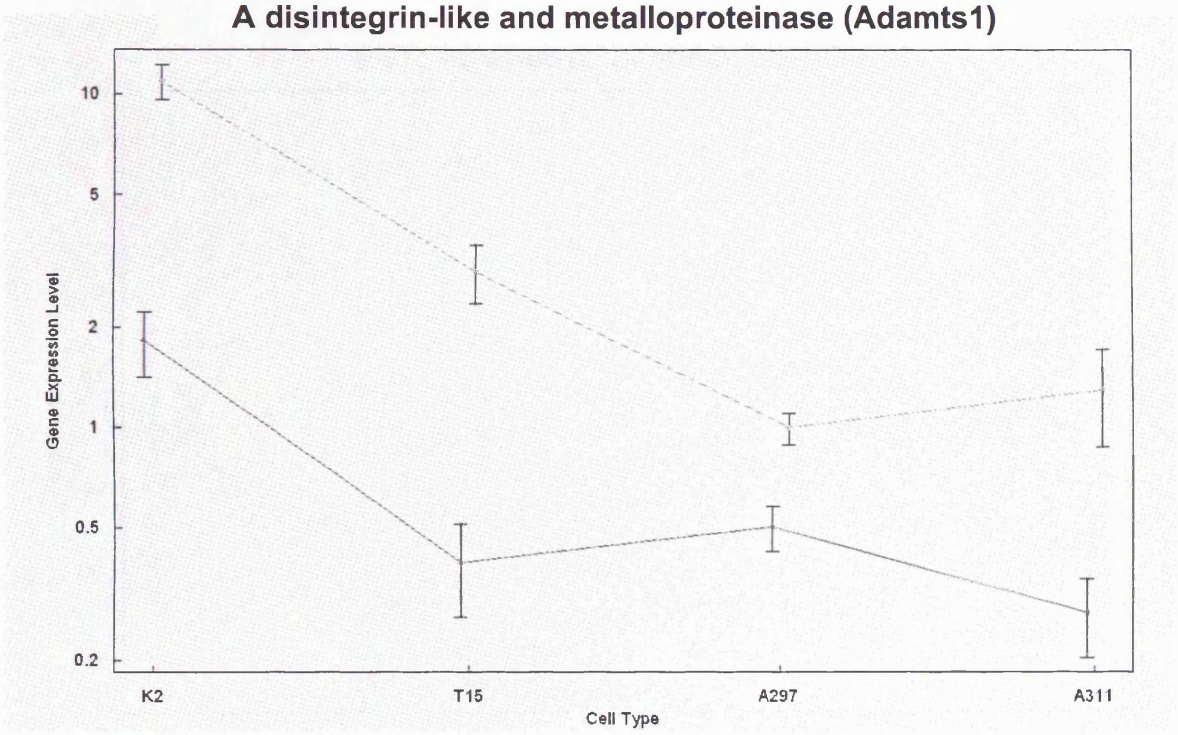

Polo-like kinase 2
Gene Expression Level
2
1
0.5
0.2
0.1
0.05
K2
T15
A297
A311
Cell Type

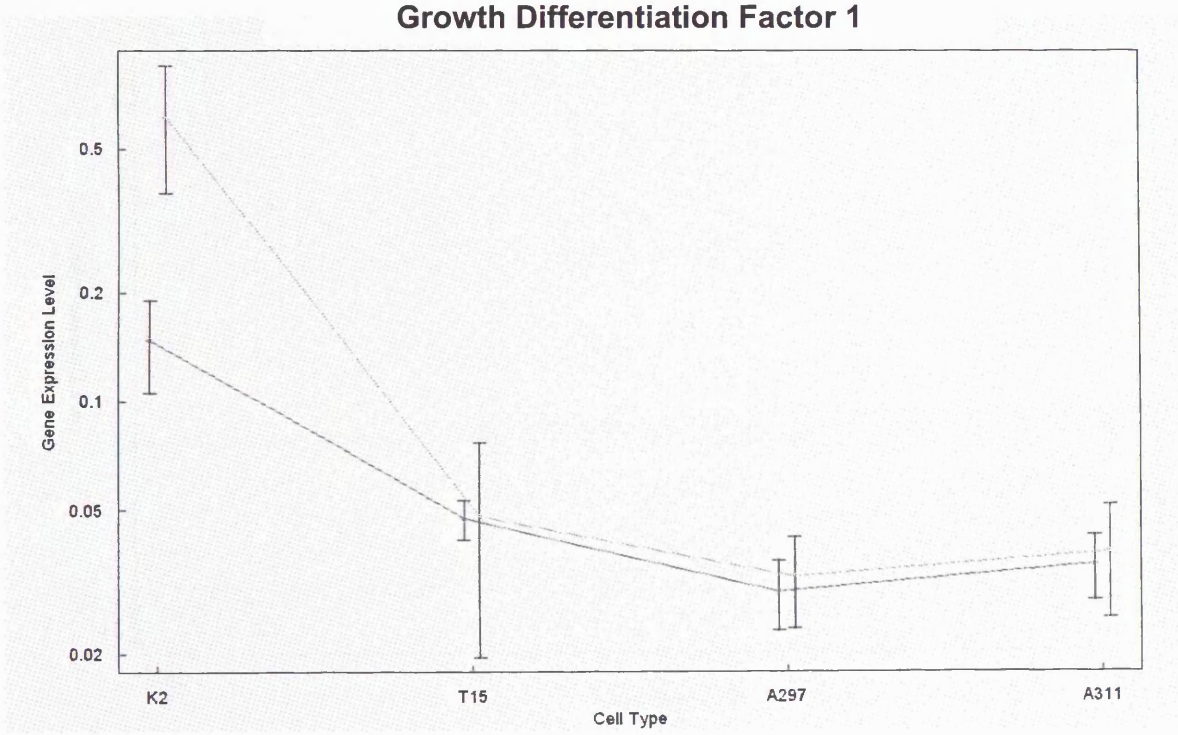

EST; Affymetrix Identifier 1375986_at
Gene Expression Level
1
0.5
0.2
0.1
K2
T15
A297
A311
Cell Type

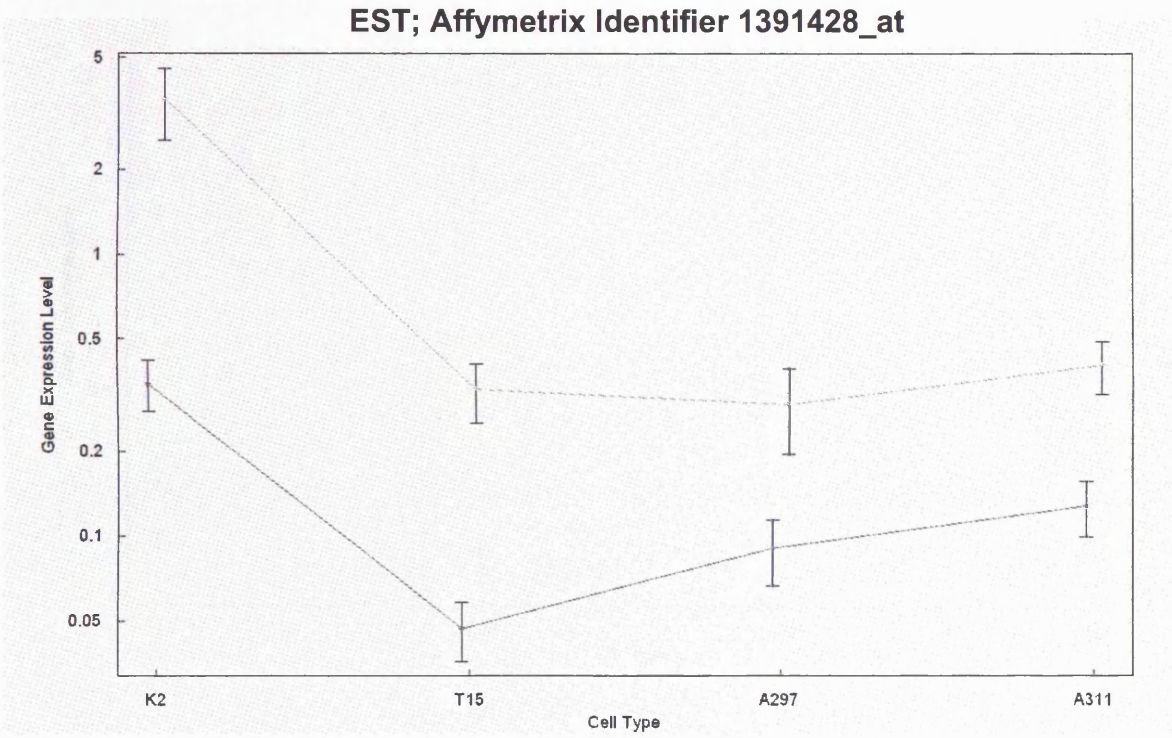

Rnase A family 4
Gene Expression Level
5
2
1
0.5
K2
T15
A297
A311
Cell Type

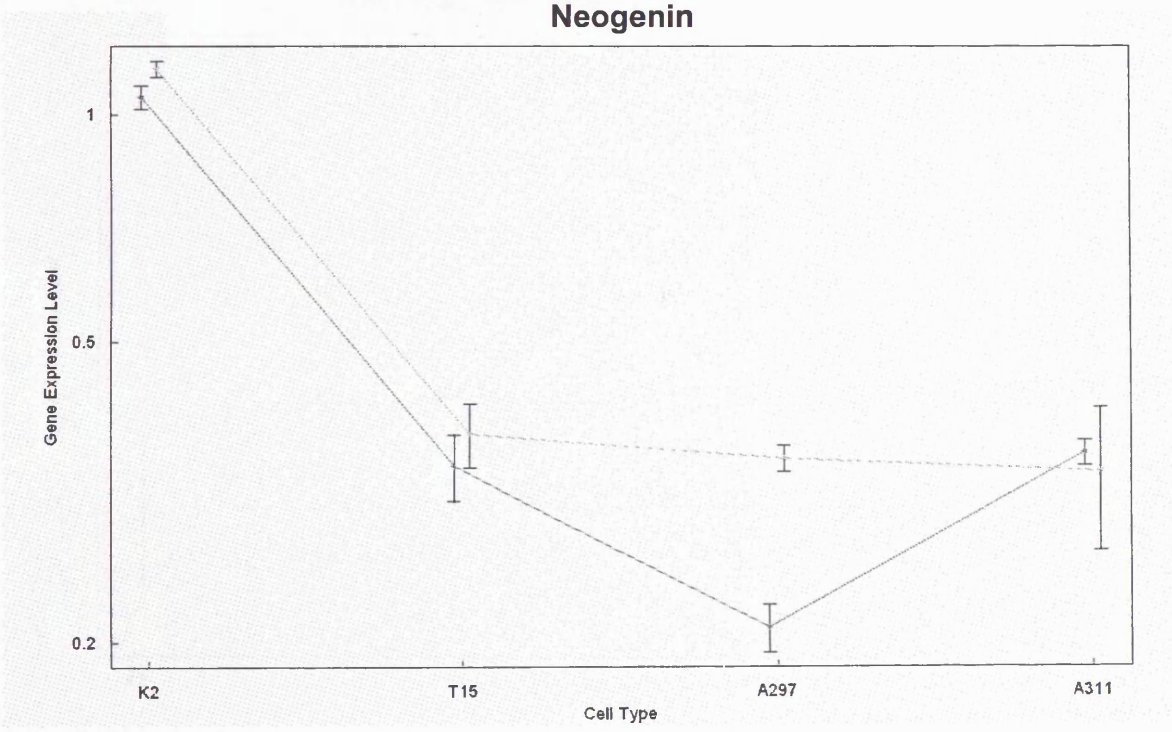

Similar to CG12279-PA
Gene Expression Level
1
0.5
0.2
0.1
K2
T15
A297
A311
Cell Type

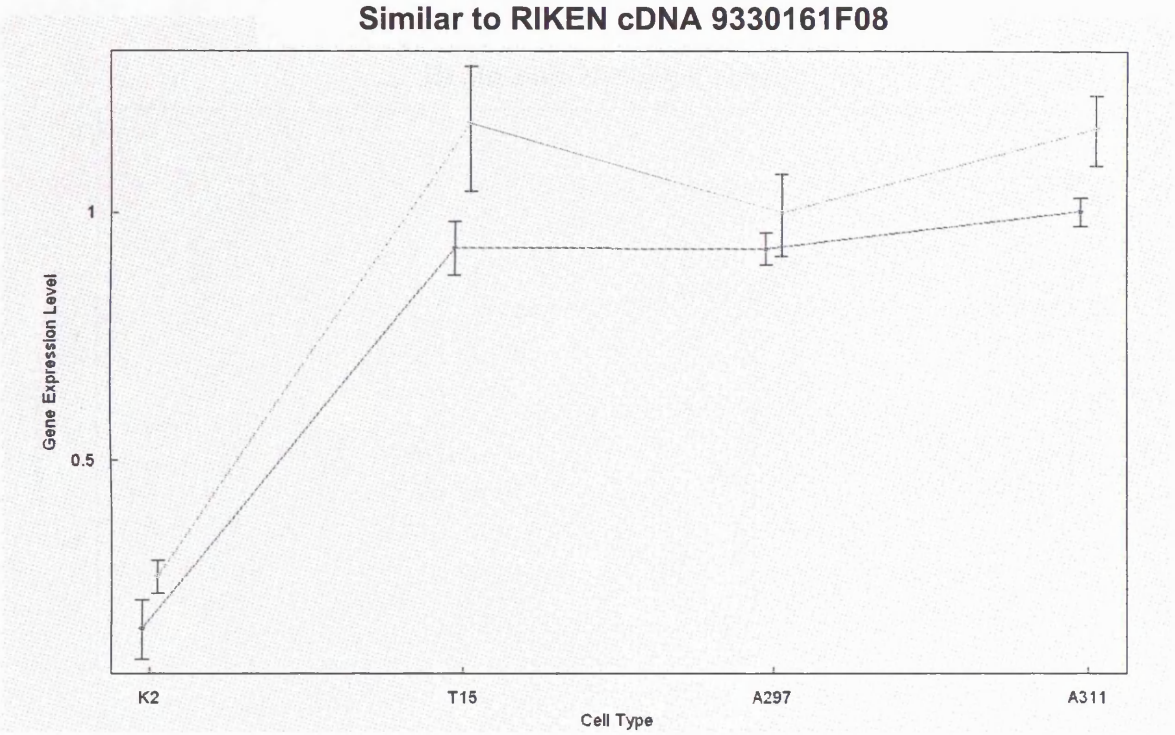

Translin-associated factor X
Gene Expression Level
1
0.5
0.2
0.1
K2
T15
A297
A311
Cell Type

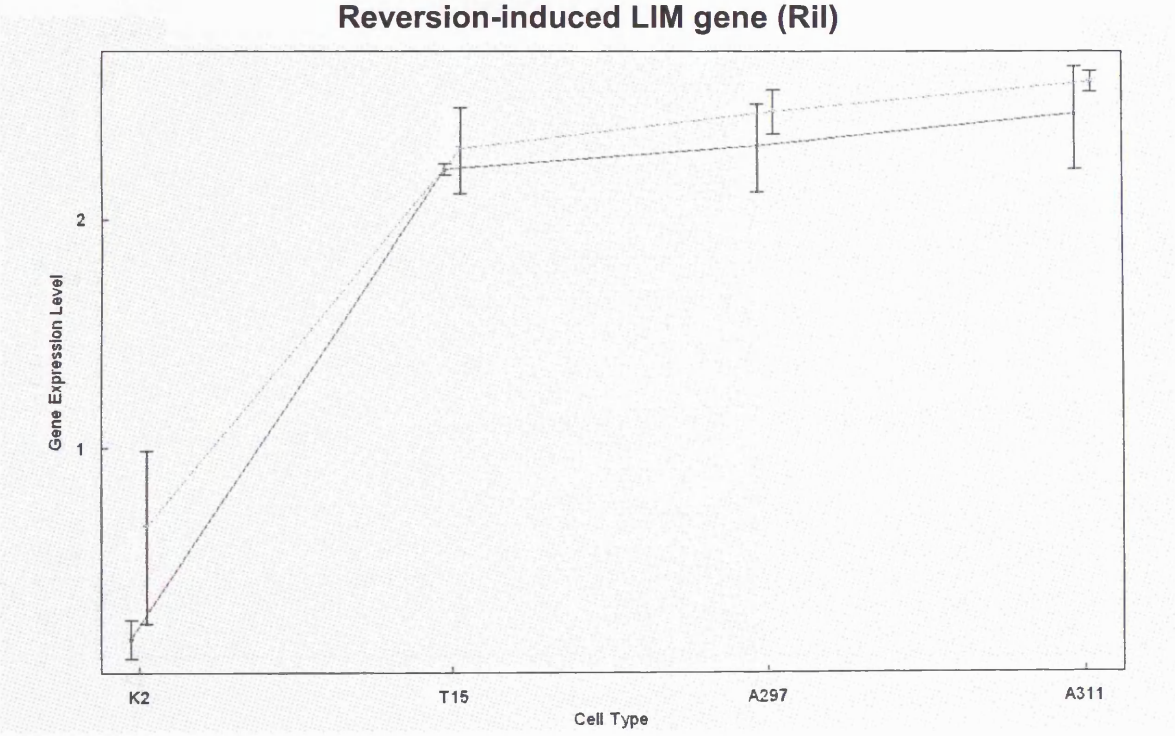

A disintegrin-like and metalloproteinase (Adamts1)
Gene Expression Level
10
5
2
1
0.5
0.2
K2
T15
A297
A311
Cell Type

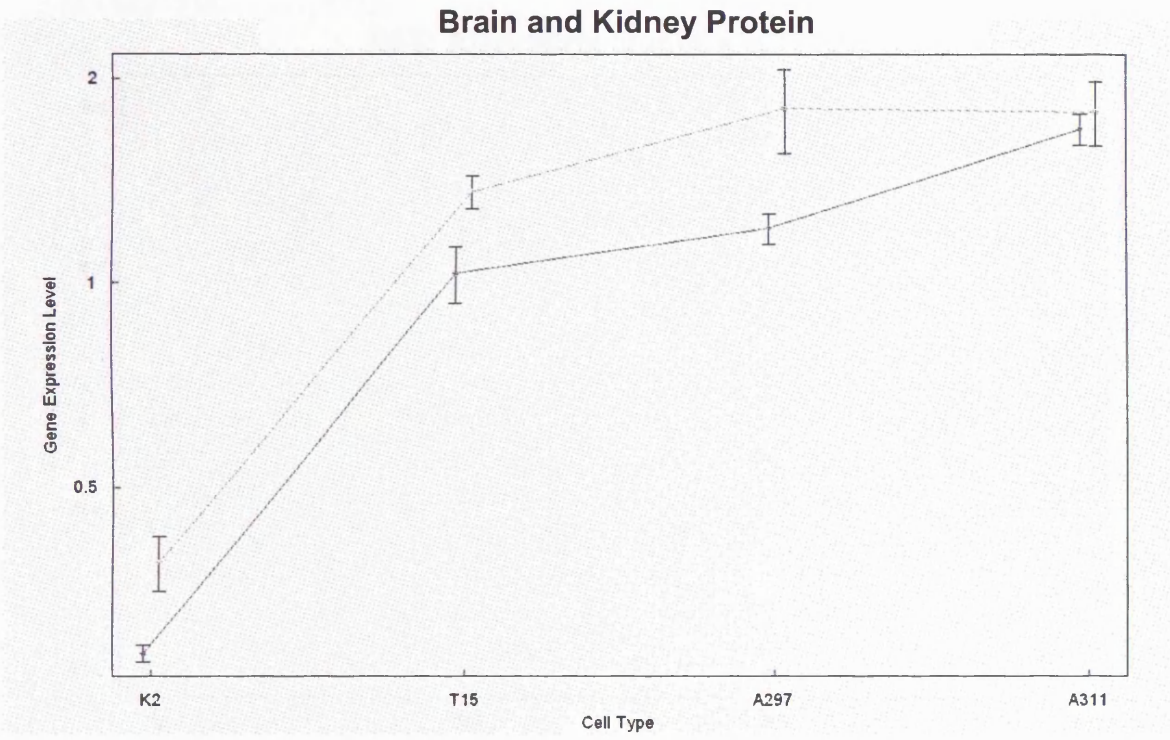

Growth Differentiation Factor 1
Gene Expression Level
0.5
0.2
0.1
0.05
0.02
K2
T15
A297
A311
Cell Type

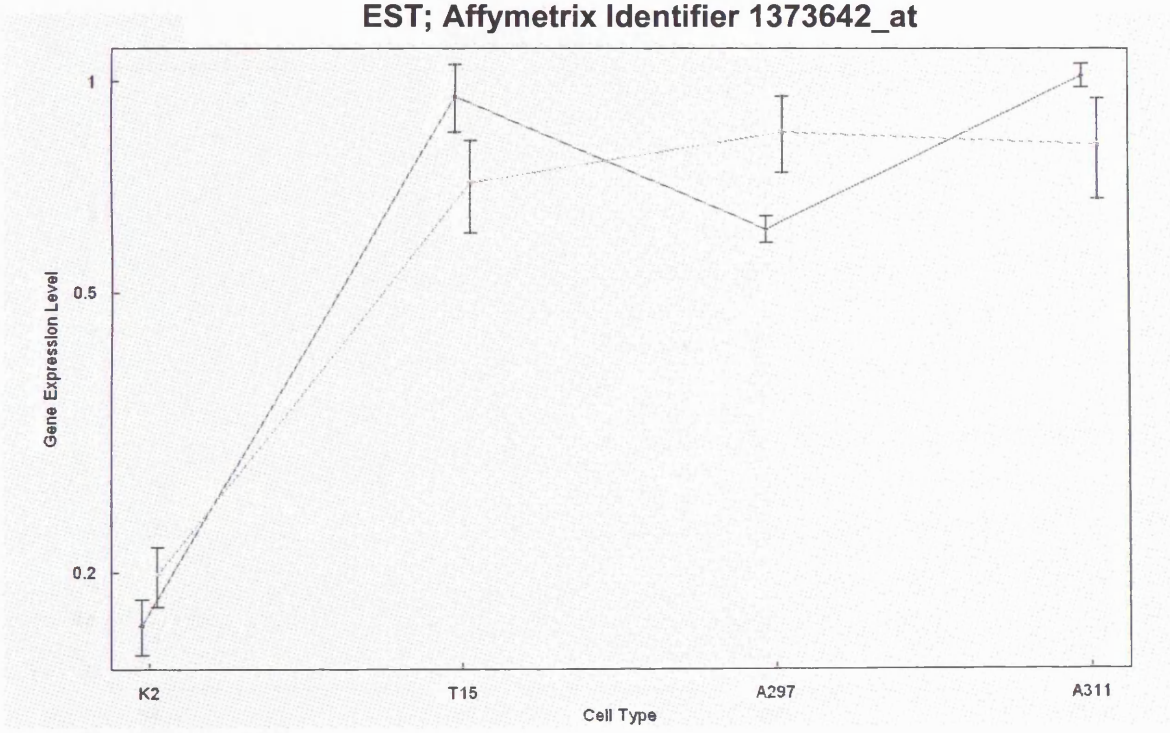

EST; Affymetrix Identifier 1391428_at
Gene Expression Level
5
2
1
0.5
0.2
0.1
0.05
K2
T15
A297
A311
Cell Type

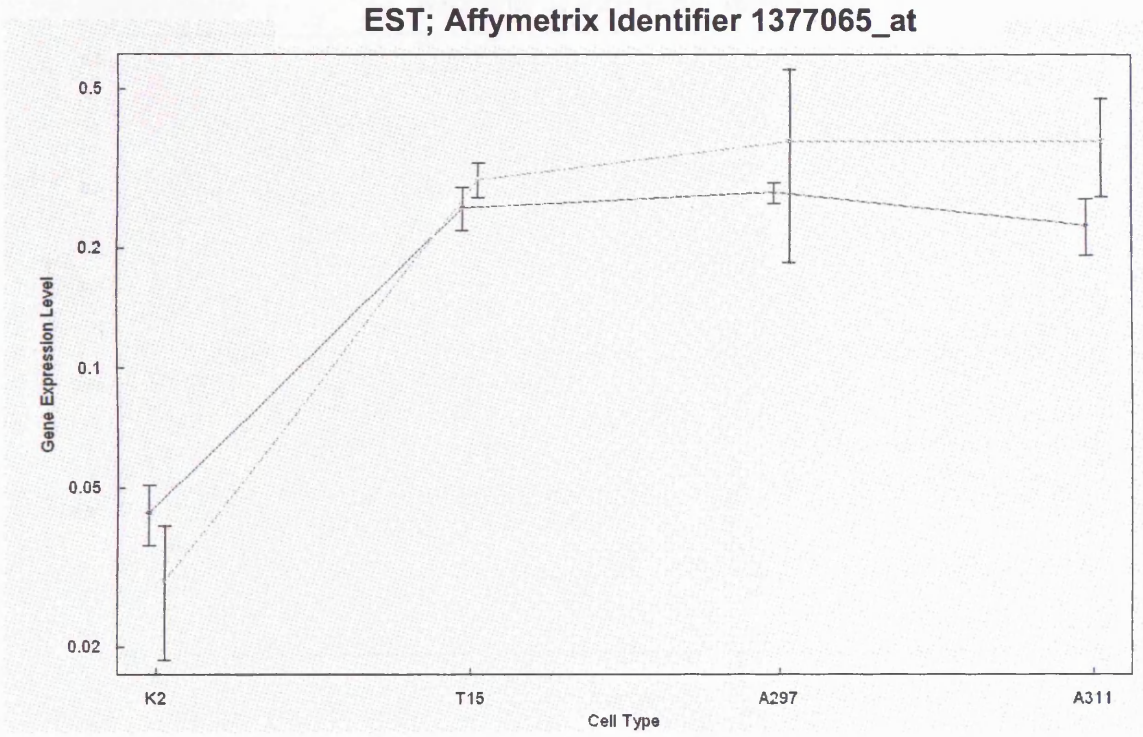

Neogenin
Gene Expression Level
1
0.5
0.2
K2
T15
A297
A311
Cell Type

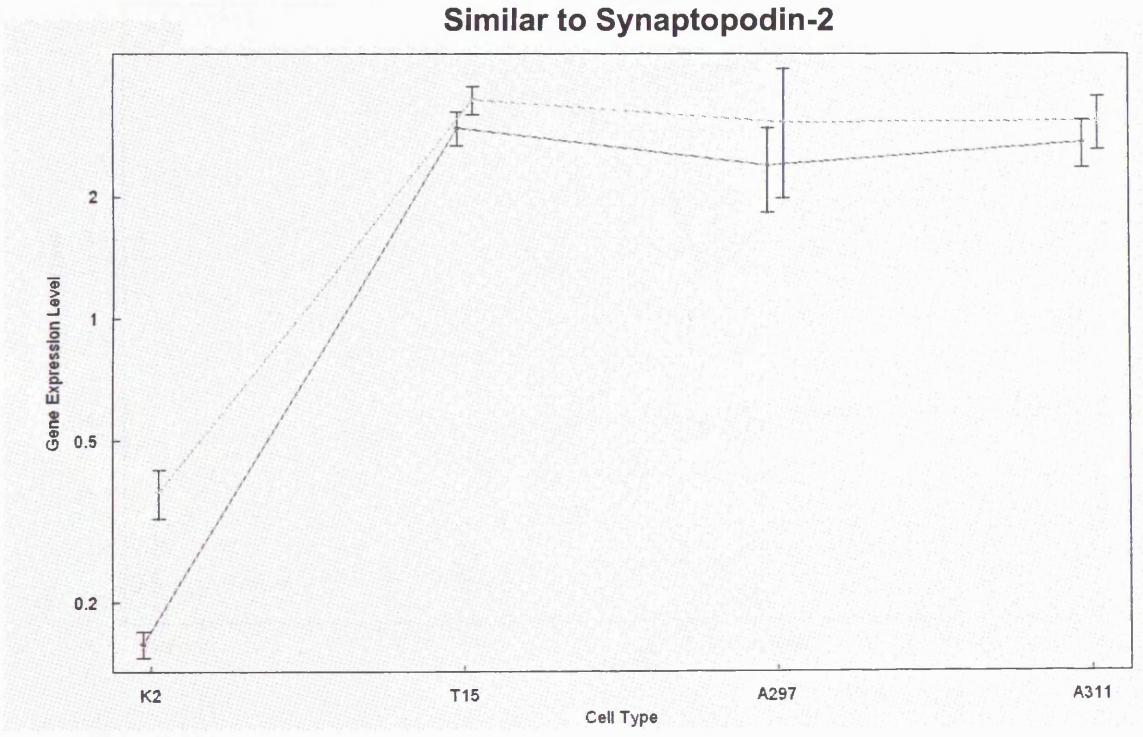

Similar to RIKEN cDNA 9330161F08
Gene Expression Level
1
0.5
K2
T15
A297
A311
Cell Type

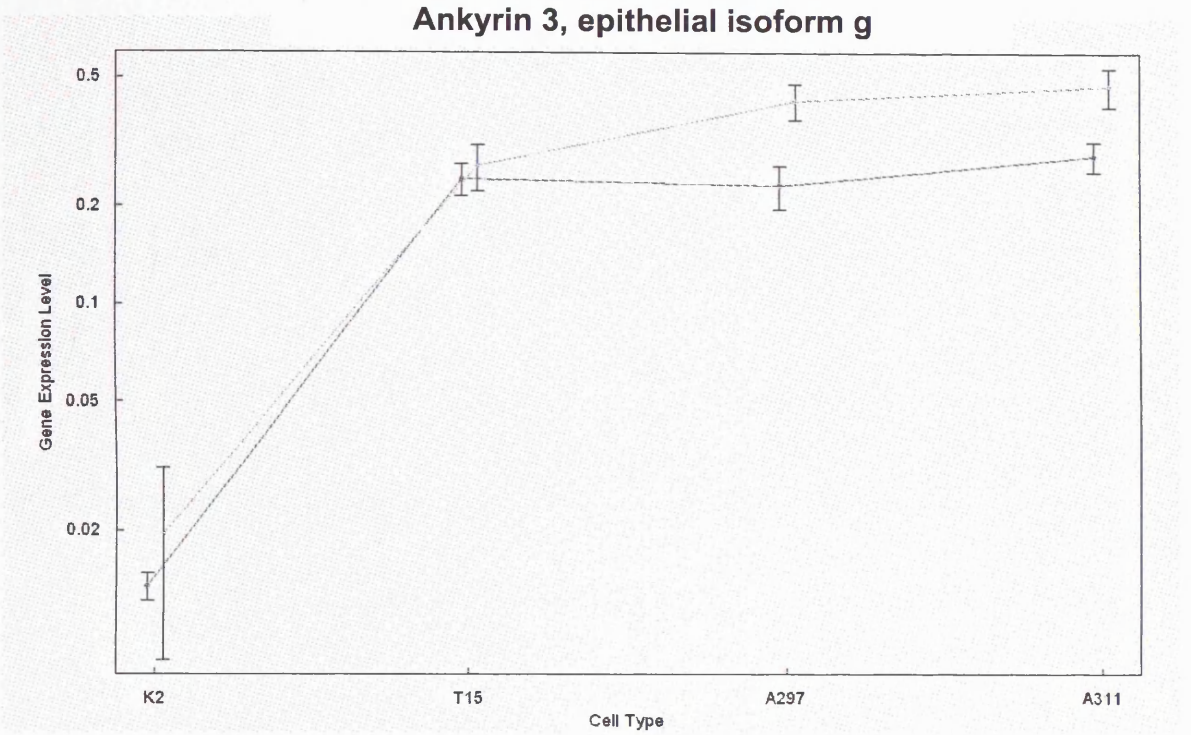

Reversion-induced LIM gene (Ril)
Gene Expression Level
2
1
K2
T15
A297
A311
Cell Type

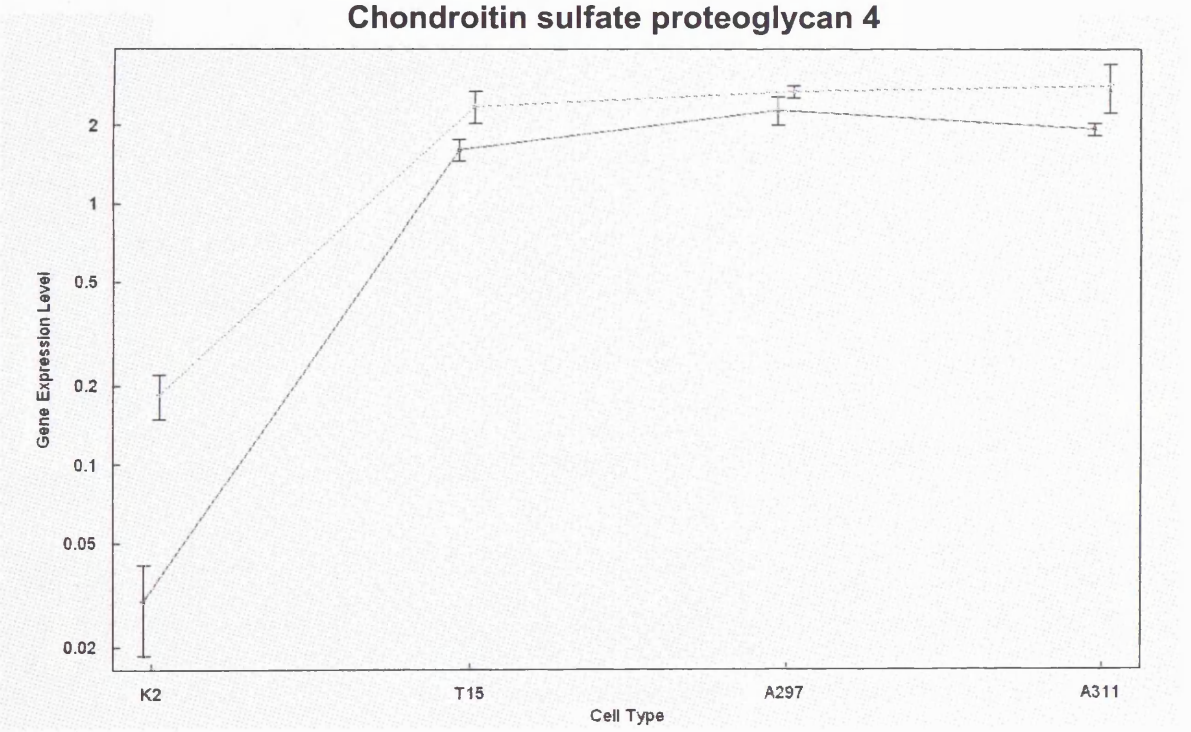

Brain and Kidney Protein
Gene Expression Level
2
1
0.5
K2
T15
A297
A311
Cell Type


EST; Affymetrix Identifier 1373642_at
Gene Expression Level
1
0.5
0.2
K2
T15
A297
A311
Cell Type


EST; Affymetrix Identifier 1377065_at
Gene Expression Level
0.5
0.2
0.1
0.05
0.02
K2
T15
A297
A311
Cell Type


Similar to Synaptopodin-2
Gene Expression Level
2
1
0.5
0.2
K2
T15
A297
A311
Cell Type


Ankyrin 3, epithelial isoform g
Gene Expression Level
0.5
0.2
0.1
0.05
0.02
K2
T15
A297
A311
Cell Type


Chondroitin sulfate proteoglycan 4
Gene Expression Level
2
1
0.5
0.2
0.1
0.05
0.02
K2
T15
A297
A311
Cell Type

Appendix 5: Author publication- Cluster analysis of networks generated through homology: automatic identification of important protein communities involved in cancer metastasis.

BMC Bioinformatics

BioMed Central

Methodology article

Open Access

Cluster analysis of networks generated through homology: automatic identification of important protein communities involved in cancer metastasis

Pall F Jonsson[1], Tamara Cavanna[2], Daniel Zicha[2] and Paul A Bates*[1]

Address: [1]Biomolecular Modelling Laboratory, Cancer Research UK London Research Institute, 44 Lincoln's Inn Fields, London WC2A 3PX, UK and [2]Light Microscopy Laboratory, Cancer Research UK London Research Institute, 44 Lincoln's Inn Fields, London WC2A 3PX, UK

Email: Pall F Jonsson - pall.jonsson@cancer.org.uk; Tamara Cavanna - tamara.cavanna@cancer.org.uk; Daniel Zicha - daniel.zicha@cancer.org.uk; Paul A Bates* - paul.bates@cancer.org.uk

* Corresponding author

Published: 06 January 2006

BMC Bioinformatics 2006, **7**:2 doi:10.1186/1471-2105-7-2

Received: 06 October 2005

Accepted: 06 January 2006

This article is available from: http://www.biomedcentral.com/1471-2105/7/2

Abstract

Background: Protein-protein interactions have traditionally been studied on a small scale, using classical biochemical methods to investigate the proteins of interest. More recently large-scale methods, such as two-hybrid screens, have been utilised to survey extensive portions of genomes. Current high-throughput approaches have a relatively high rate of errors, whereas in-depth biochemical studies are too expensive and time-consuming to be practical for extensive studies. As a result, there are gaps in our knowledge of many key biological networks, for which computational approaches are particularly suitable.

Results: We constructed networks, or 'interactomes', of putative protein-protein interactions in the rat proteome – the rat being an organism extensively used for cancer studies. This was achieved by integrating experimental protein-protein interaction data from many species and translating this data into the reference frame of the rat. The putative rat protein interactions were given confidence scores based on their homology to proteins that have been experimentally observed to interact. The confidence score was furthermore weighted according to the extent of the experimental evidence, giving a higher weight to more frequently observed interactions. The scoring function was subsequently validated and networks constructed around key proteins, identified as being highly up- or down-regulated in rat cell lines of high metastatic potential. Using clustering methods on the networks, we have identified key protein communities involved in cancer metastasis.

Conclusion: The protein network generation and subsequent network analysis used here, were shown to be useful for highlighting key proteins involved in metastasis. This approach, in conjunction with microarray expression data, can be extended to other species, thereby suggesting possible pathways around proteins of interest.

Background

Microarray experiments provide information about gene expression within the cells under study.

Expression patterns can be uncovered from large-scale microarray data by systematically grouping genes with the help of clustering methods. Co-clustering of genes can indicate that the genes in question have a similar function or that they participate in the same cellular process [1,2]. Nevertheless, microarray experiments typically yield hundreds of significantly differentially-expressed genes, making it difficult to draw biological conclusions.

Furthermore, although microarray experiments can show correlations between the expression of genes, they do not reveal the exact protein interaction mechanism.

Protein network analysis is dependent on a reliable assignment of protein-protein interactions. Protein-protein interactions are commonly studied using biochemical methods, and several different experimental methods are currently in use. Two-hybrid screens have, to date, yielded the bulk of available data [3,4]; however their level of accuracy is not particularly high and should be supported by additional evidence [5,6]. Advances in other techniques, such as tandem-affinity purification and mass spectroscopy, have also made large-scale studies increasingly feasible [7,8].

A number of computational methods, either based on sequence or structural features, have been developed to complement experimental approaches to predicting protein-protein interactions [9,10]. An increasing emphasis has been on deducing and exploring the protein-protein interaction networks that are reflected in expression data; gene networks have been inferred from gene expression data using mathematical analysis such as Bayesian regression [11-14]. Moreover, networks have been derived by complementing gene expression data with data from different sources, such as gene ontologies, phenotypic profiling and functional similarities [15-18].

Alternative techniques to network construction have also been taken, see e.g. Cabusora *et al.* [19], where a protein interaction map was created based upon the principle that interacting protein modules in one organism may be fused into a single chain in another, and Calvano *et al.* [20] who constructed the network by literature searches for information pertaining to interacting protein pairs from closely related organisms. These methods do not utilise gene expression explicitly in the network generation, rather the expression data is used as a tool to focus on the network.

Previous studies have mapped expression data of different systems onto experimentally-based networks. Ideker *et al.* [21] used gene expression changes in response to perturbation to highlight clusters within a yeast network, and Sohler *et al.* [22] made use of statistical analysis to highlight significant sub-clusters, also within a yeast network. Moreover, the dynamic aspect of yeast networks have been highlighted by de Lichtenberg and coworkers [23], who combined temporal cell cycle expression data with protein-protein interaction networks.

Here we have taken an extensive multi-genome approach, utilising a homology-based method for predicting interacting proteins [24] and further extended it by developing a scoring function, based upon sequence similarity and the amount of experimental data supporting each interaction. This scoring function has subsequently been extensively validated. In contrast to the above methodologies we go beyond data integration by considering orthologous relationships and are therefore able to create a more extensive protein interaction network – or 'interactome' – for a higher eukaryote, the rat.

In order to demonstrate the utility of our predicted interactions, expression data on tumour progression resulting in rat sarcomas with high metastatic potential were mapped onto our interactome, creating protein networks around key proteins involved in the metastatic process.

Results and Discussion

Networks of interacting proteins were constructed automatically for the entire rat (*Rattus norvegicus*) genome using the approach described in the methods section and summarised in Figure 1. The number of individual interactions was reduced from 325,087 to 151,049, when a scoring function was applied to filter out low-quality data, and was further cut down by a clustering method aimed at identifying key interconnected network nodes. The interactome data is available though the PIP (Potential Interactions of Proteins) web server [25].

Validation of the scoring function

The protein networks are composed of predicted individual interactions, each of which is assigned a score which indicates the strength of the prediction. Before examining the networks in detail it is necessary to assess the quality of the predictions and to validate the method.

Selection of cut-off value for the scoring function

Our method of constructing networks is based on homology to known interactions. It is therefore imperative to ascertain the minimum level of homology whereby the structural and functional similarity of the interacting proteins is retained. Russell *et al.* [26] have previously examined the relationship between sequence and structural divergence of interacting proteins. They found that pairs of interacting proteins can be considered structurally similar if their sequence identity is no lower than 30%. As we utilise BLAST bitscores as components for our scoring function, we tested the relationship between bitscores and sequence identity. At the 30% sequence identity level, the bitscore ranges linearly from 86–177 (see Figure 2) which, according to Equation 1, yields minimum interaction scores ranging from 9 to 10. We chose to set the minimum score for interactions at 10, to minimise possibilities of false positive results due to low homology.

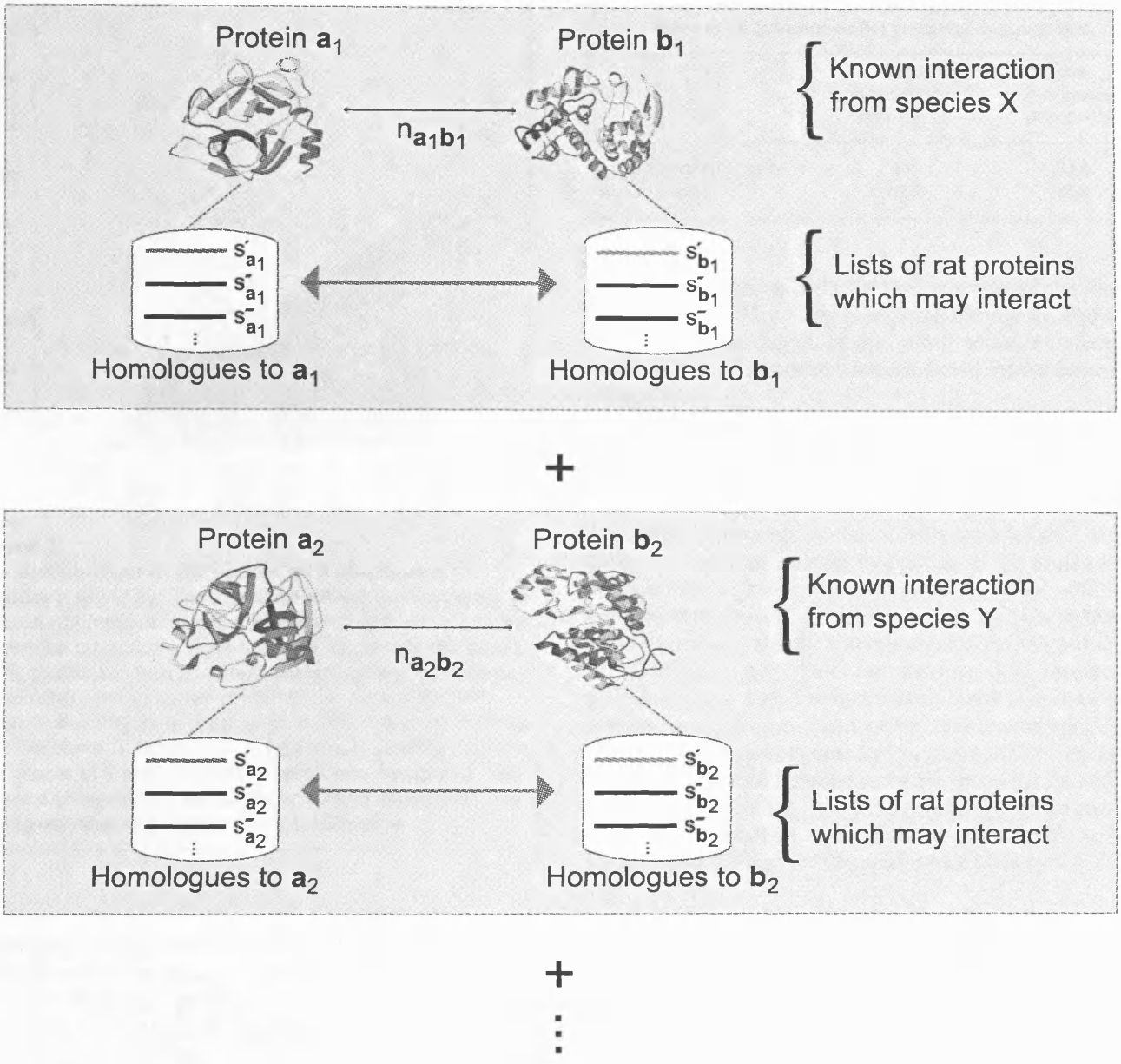


Figure 1

Inferring interactions by homology. Each interaction is inferred from homology to experimentally observed interactions. In this schematic, proteins $\mathbf{a}_1$ and $\mathbf{b}_1$ have been shown experimentally to interact in one organism, here labelled 'species X', and protein $\mathbf{a}_2$ and $\mathbf{b}_2$ in another, 'species Y'. Lists of homologues are generated for each of the proteins, ranked by their bit score (s_{a_i}, s_{b_i}, etc.). A protein from one list may interact with a protein from the other (shown by the red arrow) and potential pairwise interactions are scored according to Equation 1, based on homology to the proteins involved in the known interaction. Furthermore, interactions receive a higher score if they are derived from multiple experimental sources ($n > 1$). The score is additive, for instance, in the example here, the blue and green sequences are predicted to interact based on the interactions in 'species X' and 'species Y' and the overall score is the sum of both pairwise scores. This additive process continues over all experimentally determined protein pairs, *N*, (e.g. through 'species Z'), for which the rat sequences, labelled blue and green, are present.

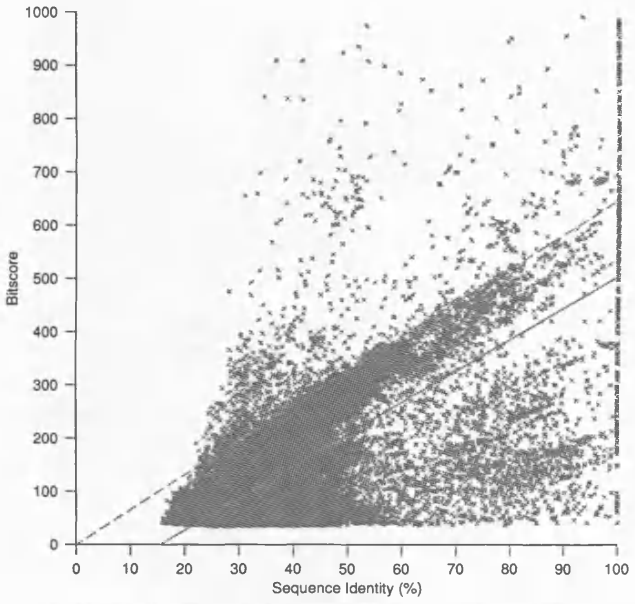


Figure 2
The distribution of bit scores as a function of sequence identity. The sequence identity and bit score of each hit when proteins in the interaction data were queried against the rat genome. The solid red line shows the best linear fit to the data and shown in dotted red is a line, starting at the origin, which contains 97% of the data in the area below it. Reading from these lines at 30% sequence identity gives bitscores of 86 and 177, respectively, yielding interaction scores of 9 and 10 when inserted into Equation 1. To ensure a stringent criteria for the minimum interaction score the higher value was selected as a cutoff score.

Identification of highly reliable interactions

Many methods for detecting protein-protein interactions can yield either false positive or false negative results, but X-ray crystal structures of complexed proteins can be considered to be a gold standard for proof. To examine the validity of our scoring function we looked at the interaction scores of rat proteins that have either been crystallised together in a complex or have a very high homology to one that has been. These scores were then compared to ones without any crystallographic evidence, i.e. those that do not interact or have not been proven to do so by crystallography.

We found that highly reliable interactions, identified by X-ray crystallography, score higher than those without crystallographic evidence, with median scores 128 and 16 respectively. This difference was significant according to a χ^2-test ($p << 0.0001$), indicating that true interactions score higher than those whose association has not been confirmed by crystallography.

Table 1: Distribution of protein-protein interaction scores. Interaction scores of X-ray crystal structures (n = 377) compared to the scores of all (genome-wide) predicted interactions.

	Percentage of interactions	
	Interaction score 0 – 10	Interaction score > 10
X-ray crystal structures	6.4	93.6
Genome-wide	43.2	56.8

Moreover, as shown in Table 1, about 94% of the interactions confirmed by X-ray crystallography score above 10, reaffirming the choice of the cutoff score, whereas just under half of all genome-wide predicted interactions score 10 and lower.

Community participation and cellular localisation

Another way of estimating the quality of the scoring function is to look at proteins participating in the same cellular process and compare them with proteins that are not thought to interact directly in a pathway. We used a clique percolation method to identify 'communities' within the network that show high interconnectivity. This yielded 37 communities of tightly interconnected proteins that will be described later. One can assume that interactions within communities are more likely to be true than interactions between communities, i.e. higher scores would be expected for intra-community interactions [27]. We found this to be true; the average score for interactions within a community was 26.2 (n = 2038) and the average score for interactions between communities was 13.5 (n = 502). This is significant at a 95% confidence level ($p = 3.1 \times 10^{-30}$).

Lastly, the protein interaction scores were examined in the context of cellular localisation. We assume that for true interactions, interacting proteins would co-localise in the same cellular compartment, at least during the time of interaction, and thus would expect predicted interactions between proteins in separate cellular compartments to be less reliable and receive a lower score. Localisation data from the Gene Ontology Consortium [28] were used, where available, for proteins within the thirty-seven protein communities. Of the protein interactions predicted, 681 (94%) were considered co-localised, with an average score of 25.8 and 41 (6%) were annotated as not sharing cellular localisation, with an average score of 13.1. The score difference is statistically significant (p = 0.001 at a 95% confidence level).

Metastatic network communities identified by cluster analysis

Metastasis is a key event that is associated with a poor prognosis in cancer patients. Metastasising cancer cells

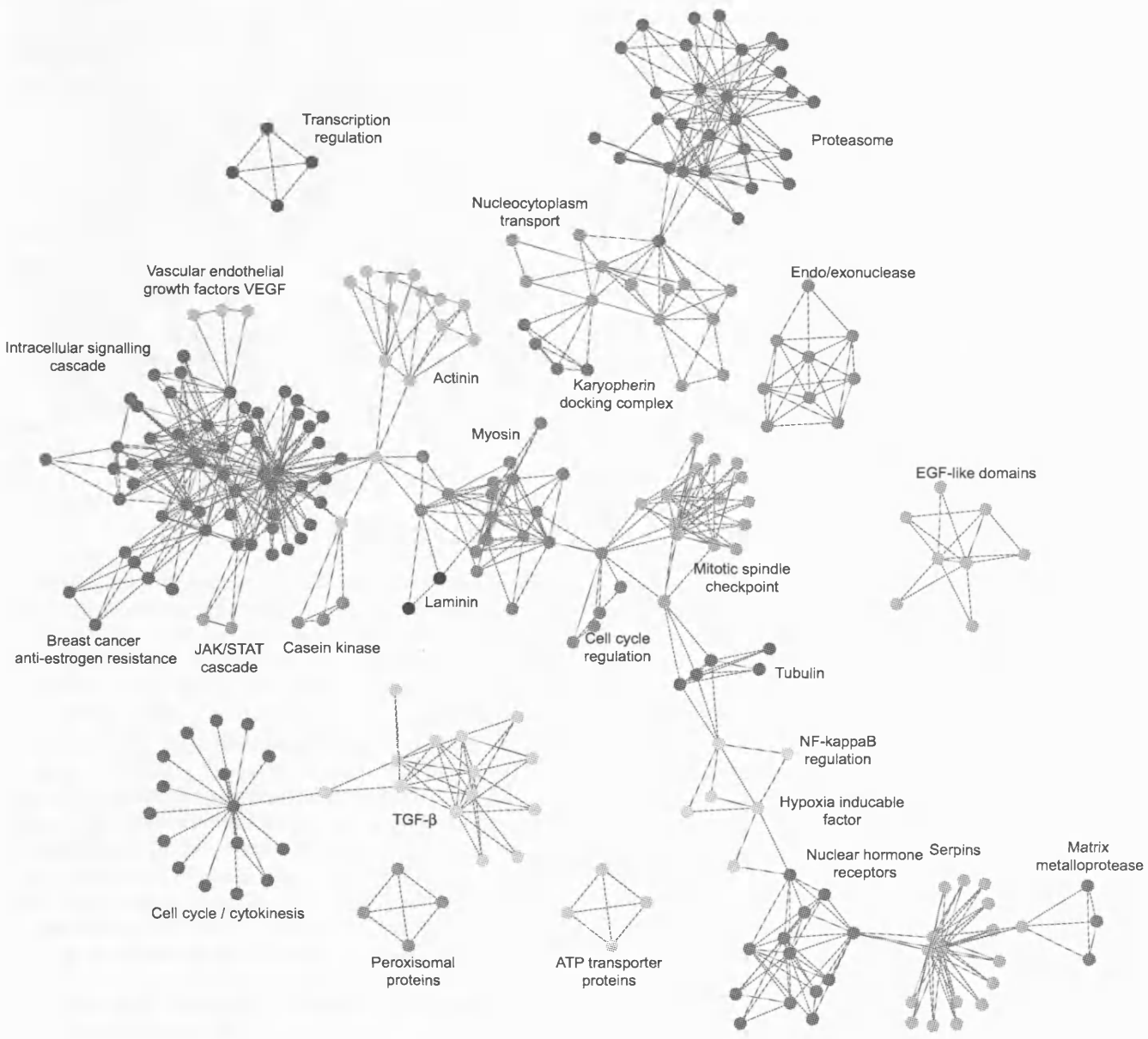


Figure 3
Identifying protein communities by cluster analysis. The communities identified by *k*-clique analysis performed on the predicted genome-wide rat protein network. The communities are distinguished by different colours and labelled by the overall function or the dominating protein class. Note that proteins, particularly at community edges, can belong to more than two communities, although this is not shown. A complete list of protein names is included as supplementary material [see Additional file 1]. The graph was created by Graphviz [61].

have the ability to break away from the primary tumour and move to different organs, making the cancer more difficult to treat. Much is unknown about the molecular biology of metastasis, but it culminates in the cancerous cells acquiring several properties, such as increased motility and invasiveness. This involves a network of cascading protein-protein interactions which have to be unravelled if an effective treatment is to be developed.

Table 2: Domain frequency within the clustered communities. The table shows the most frequently observed domains in the metastasis-related cluster communities (observed frequencies) alongside the expected domain frequencies, based on the domain composition of the whole rat genome. The *n*-fold difference was calculated from the frequency percentages (numbers within parentheses).

Domain	Observed frequency (%)	Expected frequency (%)	*n*-fold difference
Spectrin repeat	56 (6.9)	6 (0.7)	8.3
IQ calmodulin-binding motif	54 (6.6)	2 (0.2)	26.5
EGF-like domain	52 (6.4)	16 (2.0)	2.2
Protein kinase domain	47 (5.8)	12 (1.4)	3.0
SH2 domain	27 (3.3)	2 (0.3)	11.7
EF hand	25 (3.1)	7 (0.8)	2.6
Immunoglobulin domain	21 (2.6)	35 (4.3)	-0.4
SH3 domain	20 (2.4)	6 (0.7)	2.6
Calponin homology (CH) domain	13 (1.6)	2 (0.3)	5.4
Proteasome A-type and B-type	12 (1.5)	1 (0.1)	20.0
LIM domain	11 (1.3)	3 (0.4)	2.7
Transforming growth factor β-like domain	10 (1.2)	1 (0.1)	11.2

As a starting point, we used data from a microarray analysis of cell lines with different metastatic potentials (see Methods). We took the highest up- and down-regulated genes (≥ 4-fold up- or down-expression), and constructed networks around these, extending two generations from the starting point, i.e. initially including proteins that interact directly with the originating protein and then going on to include the proteins that interact with them. This subset of the rat interactome contained 10,628 interactions. We then performed a cluster analysis in order to highlight areas in the protein networks that are involved in the metastatic process. The clustering is based on a clique percolation method [29] that seeks to identify 'communities' of highly interconnected proteins that make up the essential structural units of the networks.

The community definition is based on the observation that a typical member in a community is linked to many, but not necessarily all other nodes in the community. In other words, a community can be regarded as a union of smaller, complete, fully-connected subgraphs that share nodes (see Methods section). Palla *et al.* [27] have shown that clique clustering analysis is a powerful tool to identify communities of proteins participating in the same cellular processes. Furthermore, it has been shown that subnetworks of proteins involved in a defined cellular process are more heavily interconnected by direct protein interaction than would be expected by chance [16]. Highly connected proteins are also more likely to be essential to cellular processes [30].

By applying the clustering method to our rat interactome we automatically identified 37 protein communities of highly interconnected proteins, containing 313 proteins involved in 1,094 interactions (Figure 3). The majority of the communities have been associated with cancer and metastasis. Some show a degree of overlap and are linked, the most prominent link running through the centre of the figure and containing 17 communities linked in a chain-like manner, however others are not linked, for example, the transcription regulation, which consists of only four proteins.

An initial analysis of the structural- and functional composition of the networks was performed using Domain Fishing [31], which assigns structural domains to sequences based on homology to known domains. When comparing the domain composition of the communities to domain frequencies of the whole rat genome we observed a bias towards classes of domains found in proteins involved in cytoskeletal structures, cell motility and cell-signalling (see Table 2). All but one of the most frequent domains are overrepresented when compared to the genome-wide distribution; only immunoglobulin domains appear less frequently. Spectrin repeat domains, which top the table, are found in proteins involved in cytoskeletal structure, such as spectrin, α-actinin and dystrophin. They are known to bind to calponin homology domains, which are found in both cytoskeletal and signal transduction proteins. The IQ calmodulin-binding domains work as Ca^{2+} switches for myosin which are involved in cell motility and chemotaxis. Furthermore, protein kinase domains, SH2 and SH3 domains and protein-tyrosine phosphatase participate in signal transduction and known to interact. These categories of domains,

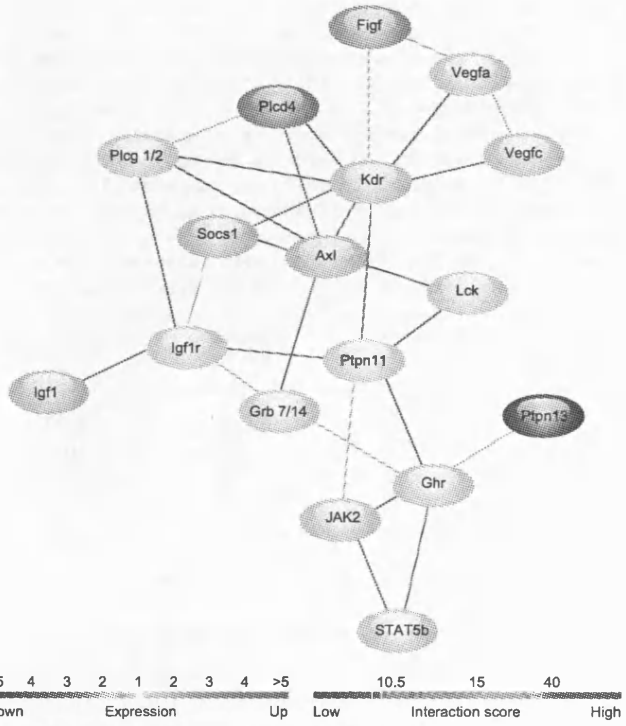


Figure 4

A closer view of a part of the 'intracellular signalling cascade'. The figure shows a subsection of the network around the intracellular signalling cascade where it extends to the VEGFs and JAK/STAT protein communities. The confidence of the interactions is shown by colour coding based on the interaction scores ranging from low-scoring blue ($10 \leq s < 10.5$) to high-scoring red ($s > 40.0$). The metastatic cell line expression levels are also shown; blue for down-regulated genes and red for up-regulated ones.

and associated functions and interactions, are all of interest in the context of cancer metastasis.

The intracellular signalling cascade

It is not the aim here to explore every member of each community, instead, with the automatic identification of metastatic-related protein communities being the primary focus, we will illustrate the value of our approach by describing a key section of the regulation pathway. The intracellular signalling cascade constitutes the head of a chain of communities (Figure 3), and as such warrants a closer investigation.

Figure 4 shows a detailed view of some of the interactions within that community, focused on the intersection with the vascular endothelial growth factors (VEGFs) and the JAK/STAT group. Many of the interactions in this network have been established either in rat or in other species; others have not been previously demonstrated and we propose that these might have a role in the context of the surrounding proteins.

Three separate groups of proteins are distinguishable: vascular endothelial growth factors (Vegfa, Vegfc, Figf) and the receptor (Kdr), which play a principle role in tumour progression and angiogenesis [32] and which have also been associated with tumour metastasis [33]; insulin-like growth factors and receptors (Igf1, Igf1r and Grb 7/14); and JAK/STAT proteins (JAK2, STATSb).

The figure shows the three ligands, Vegfa and Vegfc and Figf, at different levels of expression, all of which can bind to kinase insert domain protein receptor Kdr, a VEGF receptor, which in turn induces mitogenesis and differentation of vascular endothelial cells [34].

The interaction between Kdr and Socs1, an SH2 domain-containing suppressor of cytokine signaling 1, is plausible as Kdr has a tyrosine protein kinase domain which in a mouse homologue has been shown to interact with Socs1 [35]. Furthermore, up-regulation of Socsl has been linked with the suppression of cytokine signalling and the JAK/STAT inflammatory signalling [36-38], which is shown here further down the network; here also, Socs1 is up-regulated and JAK/STAT down-regulated.

The proposed Ptpn11-Lck interaction is based on orthology to an interaction between Ptprc and Lck in mouse. Ptpn11 and Ptprc both have tyrosine specific protein phosphatase activity. Ptpn11 is phosphorylated by tyrosine protein kinases, contains two SH2 domains and therefore could be phosphorylated by Lck.

Higher up the network are the insulin-like growth factor 1 and its receptor (Igf1 and Igf1r, respectively) which are highly implicated in different cancers [39-41]. The insulin-like growth factors are involved in several cellular processes, such as regulation of proliferation, migration, survival, size control, and differentiation [42-45]. Igf1r is overexpressed in most malignant tumours, where it functions as an anti-apoptotic agent by enhancing cell survival. Igf1 has also been shown to enhance adhesion and motility of cancer cells [46,47]; however, the exact role of Igf1r in the metastatic process has not been established. The network shown here suggests a link between the insulin-like growth factor receptor and the vascular endothelial growth factors through the highly up-regulated phospholipase delta 4 (Plcd4) and phospholipase gamma 1/2 (Plcg 1/2). The Plcg 1/2 and Igf1r interaction is based on the fact that the phospholipase has been shown to interact with insulin receptor, a close homologue of the insulin-like receptor.

Table 3: The connectivity of up- and down-regulated proteins. Observed and expected frequencies of pairwise protein interactions, categorised by their expression: N-N (non-expressed protein interacting with non-expressed protein), U-U (up-regulated protein interacting with up-regulated protein), D-D (down-regulated protein interacting with down-regulated protein) and U-D (up-regulated interacting with down-regulated). For the purpose of the classification, up-regulated proteins are those up-regulated more than 20% and down-regulated proteins down-regulated more than 20%. Expected values were calculated based on a random distribution of the expression data on the network ($p < 0.001$ for a χ^2-test).

	Observed	Expected	*n*-fold difference
N-N	8	5	1.5
U-U	121	109	1.1
D-D	17	41	0.4
U-D	71	67	1.1

Table 4: Experimental sources for building the interactome. Summary of the experiments used as a foundation for building the interactome, from most frequent (top) to least frequent (bottom). The percentage of the total is listed after each value.

Method	Frequency (%)
Two hybrid test	35,759 (69.9)
Immunoprecipitation	6,290 (12.3)
Tandem Affinity Purification (TAP)	3,503 (6.85)
Affinity chromatography	1,070 (2.09)
Copurification	572 (1.12)
Cross-linking	518 (1.01)
X-ray crystallography	511 (1.00)
In vitro binding	452 (0.88)
Biochemical/biophysical	327 (0.64)
Gel filtration chromatography	326 (0.64)
In vivo kinase activity assay	185 (0.36)
Competition binding	185 (0.36)
Immunoblotting	140 (0.27)
Cosedimentation	133 (0.26)
Gel retardation assays	106 (0.21)
Native gel electrophoresis	103 (0.20)
Other	973 (1.90)

Another distinguishing feature in the network is the highly down-regulated protein tyrosine phosphatase (Ptpn13). It has been reported that a protein tyrosine phosphatase, Ptp61F, negatively regulates the JAK/STAT pathway in *Drosophila melanogaster* [48]. Our networks suggest that the signalling protein tyrosine phosphatase, Ptpn13, may act on the JAK/STAT pathway similarly, through the dephosphorylation of the growth hormone receptor Ghr.

The few examples shown here illustrate the value of the approach in terms of revealing potential pathways and interactions that play a part in cancer metastasis, but further experimental work will be needed to confirm the validity of these predictions.

Network view of gene expression

Extracting meaningful information from microarray expression data is often difficult, especially when looking at a complex process involving a large number of genes and unknown mechanisms. Clustering of genes may be of use when trying to find genes in a common pathway and genes with related function, but it often has limitations, such as in identifying negative feedback loops [49]. Furthermore, even if key proteins are highlighted through microarray analysis, the expression data rarely reveals all proteins involved in a particular pathway.

Examining the distribution of up- and down-regulated proteins in the context of their neighbours, shows that this is indeed the case for the protein networks shown in Figure 3. The metastatic expression data was mapped onto the networks and the frequency of the up-, down- and up-/down-regulated genes interacting was examined. The results, in Table 3, indicate that if expression data from the network was randomly redistributed, the probability of observing two up-regulated proteins interacting with each other is about the same as the observed probability. That is, up-regulated proteins do not have a trend of directly interacting with each other, but are interlinked through either neutrally expressed or down-regulated proteins. Moreover, down-regulated proteins are much less likely to interact with each other than expected, demonstrating the benefit of projecting the expression data onto already built networks, as clustering similarly expressed genes and assigning to the same pathway would not be effective.

Conclusion

Expression data has previously been put into a network context, for example by incorporating gene ontology data [15] and protein interactions [50], but here we generated the networks first, mapped the expression on top, and then performed a clustering. This approach allows us to bypass some of the obstacles involved in traditional microarray analysis, such as clustering of gene expression patterns; as demonstrated here, interactions of up-up and down-down regulated genes are not necessarily co-localised. To focus on the parts of the genome-wide interaction network relevant to metastasis we first selected subnetworks around highly up- and down-regulated genes and then utilised the clique method, which highlights hubs of highly interconnected protein communities. This allows us to examine the most complex parts of the network but as a result simple linear pathways do not get included.

Table 5: Gene ontology cellular compartments. A simplified representation of gene ontology cellular compartments. Protein accessibility between compartments is represented by ones and zeros: the former indicates the possibility of interaction between respective compartments and the latter excludes any interactions.

	Extracellular	Intracellular	Cytoplasm	Nucleus	Mitochondrion	Membrane
Extracellular	1	0	0	0	0	1
Intracellular	0	1	1	1	1	1
Cytoplasm	0	1	1	0	0	1
Nucleus	0	1	0	1	0	1
Mitochondrion	0	1	0	0	1	0
Membrane	1	1	1	1	0	1

Although we believe the general approach is of value in protein network analysis, there remain some shortcomings. Most importantly, transient protein-protein interactions are unlikely to be captured by our approach. This is a direct consequence of transient not being as well documented as non-transient interactions. Moreover, the method cannot distinguish between true and false positives, for which there is limited experimental data – however, these problems will be alleviated as more high-throughput proteomic studies are completed.

The system-level approach taken here, combining information on how proteins interact with each other and how genes are expressed, is a particularly appealing way to gain understanding of complex biological processes, such as metastasis. Although not discussed here in great detail several interesting groups of interactions, at the domain level, have been highlighted as potentially important players in the metastatic process. Further dissection of these is the subject of ongoing studies and consequently to be confirmed experimentally.

This method of using homologous protein interaction data to infer protein-protein information could be particularly useful for proteins for which there is no definite binding partner information. Moreover the relative expression levels of neighbouring proteins may prove an important consideration, when protein networks are to be subsequently modulated in conjunction with disease analysis, for example by targeting the expression of a particular gene by short interfering RNA (siRNA) [51].

The approaches described in this work are readily transferable to other species and cellular processes.

Methods

Protein interaction prediction

In order to identify homologous interaction pairs for which there is experimental data, BLAST searches were run for the rat genome [52] against all proteins in the DIP [53] and MIPS Mammalian Protein-Protein Interaction databases [54].

The experimental data arises from several methods – the most frequent are listed in Table 4. The putative interactions were given confidence scores based on two factors: the level of homology to proteins found experimentally to interact, and the amount of experimental data available (see Figure 1 for an illustrationof the approach).

The score, S, was calculated for each putative interaction according to the following:

$$S = \sum_{i=1}^{N} \ln(s_{\mathbf{a}_i} s_{\mathbf{b}_i}) n, \qquad (1)$$

where $s_{\mathbf{a}_i}$ and $s_{\mathbf{b}_i}$ are sequence similarity bit scores to proteins $\mathbf{a}_i$ and $\mathbf{b}_i$, respectively, which have experimentally been shown to interact; n is the number of experiments linking protein $\mathbf{a}_i$ to protein $\mathbf{b}_i$; and N is the total number of instances where the same pair of proteins is identified as interacting through different homologues. As mentioned in the Introduction, two-hybrid experiments are prone to giving false-positive results. Although most of the interactions created here are derived through yeast two-hybrid links, it has been shown that confidence is higher for interactions detected in multiple independent yeast two-hybrid experiments [15]. This fact is reflected in the additive nature of the score, where a protein interaction that shows up repeatedly in independent two-hybrid experiments gets a higher score.

Validation

In order to test the scoring function, we created a subset of data from the RCSB Protein Data Bank [55] that specifically concentrates on stable functional protein interactions, rather than transient. Protein chains with high sequence homology to the Norwegian rat were considered

Table 6: The number of protein communities at different clustering threshold values. The number of protein communities vary as the *k*-value for clustering is changed. The table shows the total number of separate protein communities for each *k*-value.

Clustering threshold value	Number of protein communities
$k = 3$	145
$k = 4$	37
$k = 5$	12
$k = 6$	8
$k = 7$	2
$k = 8$	1
$k = 9$	1
$k = 10$	1
$k = 11$	1

($e \leq 1 \times 10^{-10}$). We distinguished biologically functional complexes (where multimeric protein chains are permanently bound and essential for the complex function) from transient ones (where protein chains may be bound to a complex but may also act as a separate functional protein on its own), by applying a method proposed by Ofran and Rost [56]. This yielded 377 binary chain interactions.

Cellular localisation of proteins was obtained from the Gene Ontology Consortium [28]. Each of the proteins identified by the cluster analysis was placed in a basic cellular localisation class as per Table 5. Protein pairs predicted to interact were considered co-localised if they were found in compatible cellular compartments.

Creation of networks around up/down-regulated genes

Rat genes that were overexpressed or underexpressed more than four-fold were used as starting points (n = 100). Networks were expanded two generations out from the starting points using protein-protein interactions whose *S*-score value was 10 or higher. The resulting 10,628 interactions were then analysed using CFinder [27], which locates maximal complete subgraphs (*k*-cliques) in the networks and then identifies 'communities' by carrying out standard component analysis of the clique-clique overlap. In this context, the variable k is defined as the number of nodes in the subgraph and a *k*-clique community is defined as the union of all *k*-cliques that can be reached from each other through a series of adjacent *k*-cliques, where cliques sharing k - 1 nodes are defined as adjacent. Table 6 shows the number of individual protein communities for different *k*-values. Thirty-seven communities were identified for k = 4, i.e. setting the subgraph size threshold to a minimum of four. Selecting the *k*-value is a balancing act; the higher the *k*-value, the smaller and more internally connected the communities become, but less connection is observed between communities. The *k*-value was selected after observing that at k = 4, reasonably large communities were formed. Proteins which shared sequence identity higher than 40% within each community, were merged together such that they appeared as a single nodes on the protein map. These merged nodes inherited all the interactions from the individual proteins before the merging process. This was done to correct for any possible redundancies caused by our homology-based method for predicting protein interactions. There was negligible change in the protein networks as a result of this.

Microarray expression data for metastatic rat cells

To investigate genes that may be important in the development of metastases, we used a rat sarcoma model in which the cell populations K2, T15, A297 and A311 have 0, 40, 90 and 100% incidence of metastasis, respectively. We performed Affymetrix microarray analysis on the four cell populations and the primary tumours that formed when the cells were injected subcutaneously into rats. All experiments were performed in triplicate, using Affymetrix rat 230A GeneChip oligonucleotide arrays [57]. Total RNA was extracted from each sample and used to prepare biotinylated target RNA; 10 μg of RNA was used to generate first-strand cDNA by using a T7-linked oligo(dT) primer. After second-strand synthesis, in vitro transcription was performed with biotinylated UTP and CTP (Enzo Diagnostics), resulting in approximately 100-fold amplification of RNA. A complete description of the procedures is included in The Paterson Institute's Affymetrix GeneChip systems protocols [58].

The target cRNA generated from each sample was processed as per the manufacturer's recommendation using an Affymetrix GeneChip Instrument System [59]. Briefly, spike controls were added to 10 μg fragmented cDNA before overnight hybridisation, arrays were washed and stained with streptavidin-phycoerythrin, and scanned on an Affymetrix GeneChip scanner. The procedure is further described in The Paterson Institute's RNA Hybridisation protocols [60]. The median fluorescence intensity value of each GeneChip was calculated and used to normalise the chips. Gene expression was considered in terms of fold-changes between non-metastatic and the median of the three metastatic samples.

Authors' contributions

PFJ constructed and analysed the protein networks and drafted the manuscript. TC carried out the microarray analysis on metastatic rat cell lines. PAB initiated the construction of the interactome and DZ its integration with the microarray data. PAB and DZ coordinated and participated in discussions and the preparation of the manuscript. All authors have read and approved the final manuscript.

Additional material

Additional File 1

Metastasis-related protein communities. List of the proteins identified by the clique analysis.
Click here for file
[http://www.biomedcentral.com/content/supplementary/1471-2105-7-2-S1.pdf]

Additional File 2

Microarray data. The microarray expression data used in this study.
Click here for file
[http://www.biomedcentral.com/content/supplementary/1471-2105-7-2-S2.pdf]

Acknowledgements

This work was funded by Cancer Research UK. The authors would like to thank members of the Biomolecular Modelling Laboratory, Ian Kerr and Holger Gerhardt at Cancer Research UK for helpful discussions.

References

1. Eisen MB, Spellman PT, Brown PO, Botstein D: **Cluster analysis and display of genome-wide expression patterns.** *P Natl Acad Sci USA* 1998, **95:**14863-14868.
2. Niehrs C, Pollet N: **Synexpression groups in eukaryotes.** *Nature* 1999, **402:**483-487.
3. Uetz P, Giot L, Cagney G, Mansfield TA, Judson RS, Knight JR, Lockshon D, Narayan V, Srinivasan M, Pochart P: **A comprehensive analysis of protein-protein interactions in Saccharomyces cerevisiae.** *Nature* 2000, **403:**623-627.
4. Ho Y, Gruhler A, Heilbut A, Bader GD, Moore L, Adams SL, Millar A, Taylor P, Bennett K, Boutilier K: **Systematic identification of protein complexes in Saccharomyces cerevisiae by mass spectrometry.** *Nature* 2002, **415:**180-183.
5. Sprinzak E, Sattath S, Margalit H: **How Reliable are Experimental Protein-Protein Interaction Data?** *J Mol Biol* 2003, **327:**919-923.
6. Bader GD, Hogue CWV: **Analyzing yeast protein-protein interaction data obtained from different sources.** *Nat Biotechnol* 2002, **20:**991-997.
7. Gavin AC, Bosche M, Krause R, Grandi P, Marzioch M, Bauer A, Schultz J, Rick JM, Michon AM, Cruciat CM: **Functional organization of the yeast proteome by systematic analysis of protein complexes.** *Nature* 2002, **415:**141-147.
8. Mann M, Hendrickson RC, Pandey A: **Analysis of proteins and proteomes by mass spectrometry.** *Annu Rev Biochem* 2001, **70:**437-473.
9. Park J, Lappe M, Teichmann SA: **Mapping Protein Family Interactions: Intramolecular and Intermolecular Protein Family Interaction Repertoires in the PDB and Yeast.** *J Mol Biol* 2001, **307:**329-938.
10. Valencia A, Pazos F: **Computational methods for the prediction of protein interactions.** *Curr Opin Struc Biol* 2002, **12:**368-373.
11. Bader JS, Chaudhuri A, Rothberg JM, Chant J: **Gaining confidence in high-throughput protein interaction networks.** *Nat Biotechnol* 2004, **22:**78-85.
12. Brazhnik P, de la Fuente A, Mendes P: **Gene networks: how to put the function in genomics.** *Trends Biotechnol* 2002, **20:**467-472.
13. Rogers S, Girolami M: **A Bayesian regression approach to the inference of regulatory networks from gene expression data.** *Bioinformatics* 2005, **21:**3131-3137.
14. Jansen R, Yu H, Greenbaum D, Kluger Y, Krogan NJ, Chung S, Emili A, Snyder M, Greenblatt JF, Gerstein M: **A Bayesian Networks Approach for Predicting Protein-Protein Interactions.** *Science* 2002, **302:**449-453.
15. Jansen R, Lan N, Qian J, Gerstein M: **Integration of genomic datasets to predict protein complexes in yeast.** *J Struct Funct Genomics* 2002, **2:**71-81.
16. Gunsalus KC, Ge H, Schetter AJ, Goldberg DS, Han JD, Hao T, Berriz GF, Bertin N, Huang J, Chuang LS, Li N, Mani R, Hyman AA, Sonnichsen B, Echeverri CJ, Roth FP, Vidal M, Piano F: **Predictive models of molecular machines involved in Caenorhabditis elegans early embryogenesis.** *Nature* 2005, **436:**861-865.
17. Lu LJ, Xia Y, Paccanaro A, Yu H, Gerstein M: **Assessing the limits of genomic data integration for predicting protein networks.** *Genome Res* 2005, **15:**945-953.
18. Rhodes DR, Tomlins SA, Varambally S, Mahavisno V, Barrette T, Kalyana-Sundaram S, Ghosh D, Pandey A, Chinnaiyan AM: **Probabilistic model of the human protein-protein interaction network.** *Nat Biotechnol* 2005, **23:**951-959.
19. Cabusora L, Sutton E, Fulmer A, Forst CV: **Differential network expression during drug and stress response.** *Bioinformatics* 2005, **21:**2898-2905.
20. Calvano SE, Xiao W, Richards DR, Felciano RM, Baker HV, Cho RJ, Chen RO, Brownstein BH, Cobb JP, Tschoeke SK, Miller-Graziano C, Moldawer LL, Mindrinos MN, Davis RW, Tompkins RG, Lowry SF: **A network-based analysis of systemic inflammation in humans.** *Nature* 2005, **437(7061):**1032-7.
21. Ideker T, Ozier O, Schwikowski B, Siegel AF: **Discovering regulatory and signalling circuits in molecular interaction networks.** *Bioinformatics* 2002, **18:**S233-S240.
22. Sohler F, Hanisch D, Zimmer R: **New methods for joint analysis of biological networks and expression data.** *Bioinformatics* 2004, **20:**1517-1521.
23. de Lichtenberg U, Jensen LJ, Brunak S, Bork P: **Dynamic complex formation during the yeast cell cycle.** *Science* 2005, **307:**724-727.
24. Goffard N, Garcia V, Iragne F, Groppi A, de Daruvar A: **IPPRED:server for proteins interactions inference.** *Bioinformatics* 2003, **19:**903-904.
25. **PIP: Potential Interactions of Proteins** [http://www.bmm.icnet.uk/~pip]
26. Aloy P, Pichaud M, Russell RB: **Protein complexes: structure prediction challenges for the 21st century.** *Curr Opin Struc Biol* 2005, **15:**15-22.
27. Palla G, Derényi I, Farkas I, Vicsek T: **Uncovering the overlapping community structure of complex networks in nature and society.** *Nature* 2005, **435:**814-818.
28. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, Harris MA, Hill DP, Issel-Tarver L, Kasarskis A, Lewis S, Matese JC, Richardson JE, Ringwald M, Rubin GM, Sherlock G: **Gene ontology: tool for the unification of biology. The Gene Ontology Consortium.** *Nat Genet* 2000, **25:**25-29.
29. Derenyi I, Palla G, Vicsek T: **Clique percolation in random networks.** *Phys Rev Lett* 2005, **94:**160202.
30. Jeong H, Mason SP, Barabasi AL, Oltvai ZN: **Lethality and centrality in protein networks.** *Nature* 2001, **411:**41-42.
31. Contreras-Moreira B, Bates PA: **Domain fishing: a first step in protein comparative modelling.** *Biomformatics* 2002, **18:**1141-1142.
32. Ferrara N, Gerber HP, LeCouter J: **The biology of VEGF and its receptors.** *Nature Med* 2003, **9:**669-676.
33. Hirakawa S, Kodama S, Kunstfeld R, Kajiya K, Brown LF, Detmar M: **VEGF-A induces tumor and sentinel lymph node lymphangiogenesis and promotes lymphatic metastasis.** *J Exp Med* 2005, **201:**1089-1099.
34. Takahashi T, Ueno H, Shibuya M: **EGF activates protein kinase C-dependent, but Ras-independent Raf-MEK-MAP kinase pathway for DNA synthesis in primary endothelial cells.** *Oncogene* 1999, **18:**2221-2230.
35. Bourette RP, De Sepulveda P, Arnaud S, Dubreuil P, Rottapel R, Mouchiroud G: **Suppressor of cytokine signaling 1 interacts with the macrophage colony-stimulating factor receptor and negatively regulates its proliferation signal.** *J Biol Chem* 2001, **276:**22133-22139.
36. Alexander WS, Hilton DJ: **The role of suppressors of cytokine signaling (SOCS) proteins in regulation of the immune response.** *Annu Rev Immunol* 2004, **22:**503-529.
37. Park EJ, Park SY, Joe EH, Jou I: **15d-PGJ2 and rosiglitazone suppress Janus kinase-STAT inflammatory signaling through induction of suppressor of cytokine signaling 1 (SOCS1) and SOCS3 in glia.** *J Biol Chem* 2003, **278:**14747-14752.

38. Ali S, Nouhi Z, Chughtai N, Ali S: **SHP-2 regulates SOCS-1-mediated Janus kinase-2 ubiquitination/degradation downstream of the prolactin receptor.** *J Biol Chem* 2003, **278:**52021-52031.
39. Furukawa M, Raffeld M, Mateo C, Sakamoto A, Moody TW, Ito T, Venzon D, Serrano J, Jensen R: **Increased expression of insulin-like growth factor I and/or its receptor in gastrinomas is associated with low curability, increased growth, and development of metastases.** *Clin Cancer Res* 2005, **11:**3233-3242.
40. Hofmann F, García-Echeverríaon C: **Blocking insulin-like growth factor-I receptor as a strategy for targeting cancer.** *Drug Discov Today* 2005, **10:**1041-1047.
41. All-Ericsson C, Girnita L, Seregard S, Bartolazzi A, Jager MJ, Larsson O: **Insulin-like growth factor-1 receptor in uveal melanoma: a predictor for metastatic disease and a potential therapeutic target.** *Invest Ophthalmol Vis Sci* 2002, **43:**1-8.
42. LeRoith D, Werner H, Beitner-Johnson D, Roberts CT: **Molecular and cellular aspects of the insulin-like growth factor I receptor.** *Endocr Rev* 1995, **16:**143-163.
43. Yenush L, White MF: **The IRS-signalling system during insulin and cytokine action.** *Bioessays* 1997, **19:**491-500.
44. Massagué J, Czech MP: **The Subunit Structures of Two Distinct Receptors for Insulin-like Growth Factors I and II and Their Relationship to the Insulin Receptor.** *J Biol Chem* 1982, **257:**5038-5045.
45. Ullrich A, Gray A, Tam AW, Yang-Feng T, Tsubokawa M, Collins C, Henzel W, Le Bon T, Kathuria S, Chen E: **Insulin-like growth factor I receptor primary structure: comparison with insulin receptor suggests structural determinants that define functional specificity.** *EMBO J* 1986, **5:**2503-2512.
46. Dunn SE, Ehrlich M, Sharp NJ, Reiss K, Solomon G, Hawkins R, Baserga R, Barrett JC: **A dominant negative mutant of the insulin-like growth factor-I receptor inhibits the adhesion, invasion, and metastasis of breast cancer.** *Cancer Res* 1998, **58:**3353-3361.
47. Andre F, Janssens B, Bruyneel E, van Roy F, Gespach C, Mareel M, Bracke M: **Alpha-catenin is required for IGF-I-induced cellular migration but not invasion in human colonic cancer cells.** *Oncogene* 2004, **23:**1177-1186.
48. Müller P, Kuttenkeuler D, Gesellchen V, Zeidler MP, Boutros M: **Identification of JAK/STAT signalling components by genome-wide RNA interference.** *Nature* 2005, **436:**871-875.
49. Armstrong NJ, van de Wiel MA: **Microarray data analysis: from hypotheses to conclusions using gene expression data.** *Cell Oncol* 2004, **26:**279-290.
50. Segal E, Wang H, Koller D: **Discovering molecular pathways from protein interaction and gene expression data.** *Bioinformatics* 2003, **19:**i264-i272.
51. Karagiannis TC, El-Osta A: **RNA interference and potential theraputic applications of short interfering RNAs.** *Cancer Gene Ther* 2005, **12:**787-795.
52. Pruitt KD, Tatusova T, Maglott DR: **NCBI Reference Sequence (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins.** *Nucleic Acids Res* 2005, **33:**D501-D504.
53. Salwinski L, Miller CS, Smith AJ, Pettit FK, Bowie JU, Eisenberg D: **The Database of Interacting Proteins: 2004 update.** *Nucleic Acids Res* 2004, **32:**D449-D451.
54. Pagel P, Kovac S, Oesterheld M, Brauner B, Dunger-Kaltenbach I, Frishman G, Montrone C, Mark P, Stumpflen V, Mewes HW, Frishman D: **The MIPS mammalian protein – protein interaction database.** *Bioinformatics* 2005, **21:**832-834.
55. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, E BP: **The Protein Data Bank.** *Nucleic Acids Res* 2000, **28:**235-242.
56. Ofran Y, Rost B: **Analysing six types of protein-protein interfaces.** *J Mol Biol* 2003, **325:**377-387.
57. **Affymetrix genechip rat expression set 230** [http://www.affymetrix.com/support/technical/datasheets/rat230_datasheet.pdf]
58. **The Paterson Institute's target preparation for Affymetrix genechip systems protocols** [http://bioinf.picr.man.ac.uk/mbcf/downloads/GeneChip_Target_Prep_Protocol-CR-UK_v2.pdf]
59. **Affymetrix expression analysis technical manual** [http://www.affymetrix.com/support/technical/manual/expression_manual.affx]
60. **The Paterson Institute's RNA hybridisation protocols** [http://bioinf.picr.man.ac.uk/mbcf/downloads/GeneChip_Hyb_Wash_Scan_Protocol-CR-UK_v2.pdf]
61. North S, Gansner E, Ellson J: **Graphviz.** 1998 [http://www.graphviz.org].

Appendix 6: Published abstract- Investigating the gene expression profiles of metastasising sarcoma cells.

The following was presented at "Cytokinematics 2004" and subsequently published in RMS Proceedings (March 2005, Volume 40/1, p37).

Investigating the gene expression profiles of metastasising sarcoma cells.

Tamara Cavanna[1], Eva Pokorná[2], Pavel Veselý[2] and Daniel Zicha[1]
[1] Light Microscopy, Cancer Research UK, London Research Institute, Lincoln's Inn Fields Laboratories, 44 Lincoln's Inn Fields, London WC2A 3PX, UK.
[2] Institute of Molecular Genetics, Academy of Sciences of the Czech Republic, Flemingovo nám. 2, 166 37 Prague 6, Czech Republic.

Metastasis occurs when cells from a primary tumour migrate to another site in the body and form a secondary tumour. The development of metastases is likely to require changes in gene expression that promote certain aspects of cell behaviour, for example motility and chemotaxis.

To investigate genes that may be important in the development of metastases, we are using sarcoma cell populations K2 and A297 derived from inbred rats. K2 sarcoma cells are spontaneously *in vitro* transformed embryonic fibroblasts, and the A297 cell population was obtained from the K2 cells by serial tumour progression. When subcutaneously injected into the rats, the K2 and A297 cells form primary tumours that shed metastases at rates of 0% and 90% respectively. Since these two cell populations have such different abilities to metastasise, we examined their behaviour and gene expression *in vitro*.

Cells were placed in the Dunn Chemotaxis Chamber with a gradient of PDGF/IGF, and phase-contrast images were acquired every 5 minutes for 16 hours. The films of K2 and A297 cells were analysed using interactive tracking software by Kinetic Imaging, and the speed and mean direction of the cells evaluated using a Mathematica notebook developed in the laboratory. We found that the K2 cells are weakly chemotactic to PDGF/IGF, whilst the A297 cells have a strong chemotactic response, and that the responses are significantly different (ANOVA $p < 0.05$). We

also found that the K2 cells move at a speed of 11 μm per hour, whereas the A297 cells move at 17 μm per hour. The difference in speed between the two populations is significant (ANOVA $P < 0.01$).

In order to identify genes that may contribute to the observed *in vivo* malignancy and *in vitro* motility, we compared the gene expression of the two cell populations using the Affymetrix system. We found that 297 genes were significantly differentially expressed, and selected 10 of these for further study. We have validated the differential expression of some of these genes by RT-PCR and western blotting, and are now in the process of investigating their function in cell motility and other behaviour that may be relevant to metastasis.